Models for the relative biological effectiveness in proton therapy: on the impact of cell-irradiation design

Odin Nødset Alvestad

Supervisors: Sara Pilskog, Kristian Ytre-Hauge and Eivind Rørvik

Master thesis in medical physics and technology

Department of Physics and Technology

University of Bergen

June 2020

Acknowledgements

Firstly, I would like to express my endless gratitude to my supervisors, Dr. Sara Pilskog, Dr. Kristian Ytre-Hauge, and Dr. Eivind Rørvik. Thank you for your invaluable guidance and dedication throughout this project.

Thank you, Sara, for your close and thorough guidance and numeral proofreadings. Your dedication has been essential for the existence of this thesis. Thank you, Kristian, for providing several of the needed components for this thesis and introducing me to the interesting field of proton therapy. Thank you, Eivind, for providing me with a fantastic basis for this project with the work in your Ph.D.

I would also like to thank Ph.D. candidates Lars Fredrik Fjæra, and Helge Henjum for providing the scripts, and helping with the technicalities during the simulation process in this thesis.

Thank you to all my friends and fellow students at the institute, for making my time as a student both academically and socially rewarding. Special thanks to Simen and Alex for sticking with me from day one at the university.

Finally, I want to thank my mother, father, and brothers for always supporting me in the stressing times during this project.

Bergen, June 2020

Odin Alvestad

Abstract

Background: The use of protons was first suggested for use in treating cancer in 1946, due to its preferable dose distribution. In clinical practice, a constant Relative Biological Effectiveness (RBE) of 1.1 is used in treatment planning for protons, to account for the difference in biological effects between photon and proton irradiation. The RBE is, however, known to vary with variables such as physical dose levels, tissue type and Linear Energy Transfer (LET). RBE models are generally based on data derived from *in vitro* experiments.

The aim of this study was to investigate how the restrictions of input data affect the estimates of RBE for protons predicted by such models.

Methods: A database containing 98 *in vitro* experiments was gathered for this thesis. The database was divided into smaller homogenous databases, using the SPSS Two Step clustering algorithm. These databases were fitted to common equations for RBE_{min} and RBE_{max}. Finally, a Monte Carlo (MC) simulation was performed in order to recalculate the dose response in a spread out Bragg peak (SOBP), delivered to a planning target volume (PTV), predicted using the obtained RBE limits.

Results: The cluster analysis divided our database into five clusters of different sizes. The fitting process provided fitting variables in the range of -0.031 to 0.097 $\left(\frac{keV}{\mu m}\right)^{-1}$ for RBE_{min}, and -0.006 to 1.271 $Gy\left(\frac{keV}{\mu m}\right)^{-1}$ for RBE_{max}. Furthermore, the recalculations predicted doses in the range of 0.60 to 3.50 Gy(RBE) to the PTV, whereas the highest dose values were located in the distal edge of the SOBP.

Conclusion: Databases overrepresented by high α values, thus high $\left(\frac{\alpha}{\beta}\right)_x$ values, seemed to predict the highest doses. High D_{min} databases also predicted high doses. Databases representing low LET_d, and low α values, predicted the lowest doses in this study.

Contents

ACKNOWLEDGEMENTS ..... **II**

ABSTRACT ..... **IV**

CONTENTS ..... **VI**

1. INTRODUCTION **1**

1.1 RADIOTHERAPY (PHOTON THERAPY) 2

1.2 PROTON THERAPY 3

1.3 PROJECT OBJECTIVES/MOTIVATION 4

2. THEORY OF PARTICLE THERAPY **5**

2.1 INTERACTIONS WITH MATTER 5

2.1.1 Photon interaction with matter *5*

2.1.2 Proton interaction with matter *6*

2.1.3 Dosimetry *8*

2.1.4 Depth-Dose profile *8*

2.1.5 Stopping power *10*

2.1.6 Linear Energy Transfer *12*

2.2 RADIOBIOLOGY 13

2.2.1 Colonial cell experiments *13*

2.2.2 The Linear-Quadratic model *15*

2.2.3 Relative Biological Effectiveness *16*

2.3 CLUSTER ANALYSIS 19

3. MATERIAL AND METHODS **21**

3.1 THE DATABASE 21

3.2 CLUSTER ANALYSIS 22

3.3 MODEL COMPARISON 27

3.4 RBE MODELLING/ESTIMATION 27

3.5 FLUKA SIMULATION AND DOSE CALCULATIONS 28

4. RESULTS 31

4.1 DATABASE CALCULATIONS 31

4.2 CLUSTER ANALYSIS 34

4.3 MODEL COMPARISON 41

4.3.1 McNamara [81] (15%) 41

4.3.2 Mairani [82] (42%) 42

4.3.3 Wilkens and Oelfke [28] (42%) 42

4.3.4 Peeler [83] (50%) 43

4.3.5 Wedenberg [79] (54%) 43

4.3.6 Rørvik [25] (71%) 43

4.3.7 Tilly [67] (81%) 44

4.4 RBE MODELLING 44

4.5 FLUKA SIMULATION AND DOSE CALCULATONS 52

5. DISCUSSION 56

5.1 DATABASE 57

5.1.1 Quality of the cluster analysis 57

5.1.2 Clustering results 58

5.2 RBE MODELING 60

5.3 FLUKA SIMULATION AND DOSE CALCULATIONS 61

5.4 FURTHER WORK 62

6. CONCLUSION **63**

BIBLIOGRAPHY **64**

APPENDIX A **69**

APPENDIX B **70**

APPENDIX C **74**

APPENDIX D **78**

APPENDIX E **79**

List of figures:

Figure 1: Linear attenuation coefficient 6
Figure 2: Main processes protons interacts with matter 8
Figure 3: Depth dose profiles 10
Figure 4: Dose average LET depth distributions 13
Figure 5: Example of cell survival curves 14
Figure 6: Interpretation of the LQ model 16
Figure 7: Depth dose distribution of several RBE models 19
Figure 8: Loss of information and Silhouette scores 35
Figure 9: Predictor Importance of input variables 36
Figure 10: Cluster comparison of input variables 38
Figure 11: Cluster comparison of evaluation fields 39
Figure 12: $RBEmin$ fit to the experimental data. 46
Figure 13: $RBEmax$ fit to the experimental data 48
Figure 14: Represntation of “goodness of fit” for Cluster 2 nand 4. 49
Figure 15: One dimensional dose distribution 52
Figure 16: Two dimensional dose distribution 53
Figure 17: DVH of biological dose. 54
Figure 18: Two-dimensional RBE distribution. 55

Appendix:

Figure 19: Loss of information and Silhouette scores (old) 71
Figure 20: Predictor Importance for input variables (old) 71
Figure 21: Cluster comparison of input variables (old) 72
Figure 22: Cluster comparison of evaluation fields (old) 73
Figure 23: $LETd$ distribution 79

List of Tables:

Table 1: Description, value, and unit of *Bethe-Bloch* variables. 11
Table 2: The data gathered in this thesis 32
Table 3: Mean and standard deviations for each cluster model 37
Table 4: Fitting variable for RBEmin 45
Table 5: Fitting variable for RBEmax. 46
Table 6: D(0%), D(50%), and D(100%) for the PTV 54

Appendix:

Table 7: Values used to normalize LETd 69
Table 8: Mean and standard deviation for each cluster model (old) 70
Table 9: Description of the scripts used simulation process. 78

List of abbreviations

AIC:	Akaike's Information Criterion
BIC:	Schwarz's Bayesian Information Criterion
BP:	Bragg Peak
CHO:	Chinese Hamster Ovary (Cells)
DNA:	DeoxyriboNucleic Acid
DSB:	Double Strand Break
DVH:	Dose Volume Histogram
Gy:	Gray [Unit of Dose]
IndRep:	Induced Repair (model)
LET:	Linear Energy Transfer
LET_d:	Dose averaged Linear Energy Transfer
LQ:	Linear Quadratic (model)
LQC:	Linear Quadratic Cubic (model)
MC:	Monte Carlo
NTCP:	Normal Tissue Complication Probability
PIDE:	Particle Irradiation Data Ensemble
PTV:	Planning Target Volume
RBE:	Relative Biological Effectiveness
RT:	Radiation Therapy
SOBP:	Spread Out Bragg Peak
SSB:	Single Strand Break
TCP:	Tumour Control Probability

1. Introduction

As of 2018 it was estimated, using the global cancer database GLOBOCAN 2018, that there would be 18.1 million new cases of cancer, and 9.6 million cancer deaths worldwide in 2018 [1]. The risk of developing cancer is, according to the American Cancer Society, 40% for men and 38% for women. The following risk of dying as a result of cancer is for men 22% and for women 19%, meaning that statistically about one out of every five people will die as a result of cancer [2].

In addition to surgery and chemotherapy, radiotherapy is one of the main modalities for treatment of cancer. External beam radiotherapy can be divided into two main categories: photon therapy and particle therapy. Photon therapy makes use of high energy X-rays, i.e. photons, while particle therapy utilizes massive particles such as protons and carbon ions to kill the tumour cells.

The goal of radiotherapy is to kill or sterilise the tumour cells using ionizing radiation, while, as far as possible, sparing the healthy tissue surrounding the tumour. Cells are mainly killed or inactivated through damage to the DeoxyriboNucleic Acid (DNA) in form of single strand breaks (SSB) or double strand breaks (DSB) to the helical structure of the DNA. Protons are, in general, better than photons at producing non-reparable damage to the DNA structure.

The interest of radiotherapy using protons has recently spiked in Norway as two proton centres are under planning, one in Bergen and one in Oslo. Treatment with protons is planned to start in 2024 [3].

1.1 Radiotherapy (Photon therapy)

In 1895, the German physicist Wilhelm Conrad Röntgen discovered the phenomenon which he later termed X-rays, and its use as treatment for breast cancer was put in action by Emil Herman Grubbe only a year after this discovery [4]. Following the discovery, radiotherapy has increasingly been used to treat cancerous diseases either by itself or in combination with surgery or other medicinal treatments.

Radiotherapy (RT), also known as radiation therapy, is treatment primarily used to treat cancer. The goal of radiotherapy is to irradiate the tumour using massive particles, such as protons or electrons, or high-energy photons. Ideally killing the tumour cells, while sparing healthy tissue.

Sparing healthy tissue completely is not feasible, and side effects will therefore always be a problem. These side effects differ from person to person and depend on several factors, such as the type of cancer treated, the delivered dose, the general health of the patient and the location of the tumour relative to healthy organs. Examples of side effects from radiotherapy can be skin problems in the form of e.g. itching and blistering, fatigue and more severely, organ failure or the development of a secondary cancer. More location specific side effects can be nausea, hair loss and tooth decay in the head and neck region, difficulties swallowing and shortness of breath in the chest region, and diarrhoea and rectal bleeding in the pelvis region [5, 6].

An important aspect of RT research is to increase the so-called “therapeutic ratio”, also known as the “therapeutic window”. This is the ratio of the probability of controlling, or eradicating, cancerous cells in a patient, called the tumour control probability (TCP), and the probability of harming normal tissue in the process, called the normal tissue complication probability (NTCP). Technological advances, such as applying advanced imaging techniques to cover the intended area more precisely, or the implementation of fractionated dose deliveries to allow more recovery time for healthy tissue, have helped increasing the therapeutic ratio. Another modality that can help increase the

therapeutic ratio is the use of proton beams to kill the tumour, as a proton beam has an advantageous dose distribution [7].

1.2 Proton therapy

The idea of using protons to treat cancer was first suggested by Robert Wilson in 1946, with the main motivation being the protons' preferable dose distribution. The protons deposit a low dose in the proximal region of the beam path, followed by a high rise in dose deposited to a peak, before consecutively falling to zero. This peak in energy deposit is called the Bragg Peak (BP). The BP is advantageous, because it means a lower integral dose is delivered to healthy tissue compared to photon based therapies, and also the fact that the dose deposition falls to zero after the BP make it advantageous close to vital Organs at Risk (OARs).

It is generally agreed that the biological effects of photon, and proton radiation is different, even at the same dose levels. To account for this difference, a scaling factor called the Relative Biological Effectiveness (RBE) is used. Clinically a constant RBE of 1.1 is used, even though cell survival data shows that this is not always correct, especially for high linear energy transfer (LET) values. RBE models have been created, that uses the knowledge we have from photons to predict the effect of protons, to account for the effects of a variable RBE. Most of these models are phenomenological, meaning that they are based on empirical data through some sort of regression.

As a result, the same described dose could be given to the tumour with particle therapy, while the dose to healthy tissue would be significantly lower compared to the deposition from photon therapy. This would theoretically also result in less severe side effects.

1.3 Project objectives/motivation

Cell survival experiments where cells are irradiated at different dose levels, and different LET values, form the basis for modelling of the RBE of protons, representing the increased effect of protons compared to the standard radiotherapy with photons. Previously performed cell experiments vary in the range of doses the cells were exposed to, types of cells irradiated, and experimental set-up, influencing e.g. the LET distribution.

RBE models are generally derived by fitting experimental data, derived from cell survival experiments, to an equation using some form of regression algorithm. Research has, however, suggested that the upper and lower extreme RBE boundaries implemented in these models are dependent on the varying ranges of input data included in their database.

In this thesis, we wished to further research the effects restrictions on the database have on the model estimations of the RBE. We wish to use cluster analysis to group similar cell survival experiments to investigate any trends in the RBE boundaries, RBE_{min} and RBE_{max}, as these previously have been shown to be dependent on their input database [8].

A sizable database of experimental cell survival data was gathered for this study. The database was divided into smaller homogenous databases using a cluster analysis algorithm. These smaller databases were further fitted to common RBE_{min} and RBE_{max} definitions, used in RBE models. These refits were then compared to each other, and existing RBE models, in order to identify how the input data of the database affected the biological response.

2. Theory of particle therapy

2.1 Interactions with matter

A substantial part of this thesis is based on comparison of photon and proton radiation, therefore a short introduction to how these particles interact with matter is provided in this chapter.

2.1.1 Photon interaction with matter

Photons are massless, neutrally charged packs of energy. As photons traverse matter, they can interact to produce free electrons by three main processes: the photoelectric effect, Compton scattering and pair production [9, 10]. Through the photoelectric effect, the photon is absorbed by an atomic orbital electron, which is in turn ejected from the atom. In Compton scattering, or incoherent scattering, the photon interacts with an atomic electron as if it were “free”, meaning that the electron binding energy is significantly lower than the energy of the incident photon. Further the photon hits the “free” electron, passing on a portion of its energy and emitting the electron at an angle. Pair production can only occur if the incident photon has an energy of at least twice the rest mass of an electron ($1.022\ MeV$). The photon interacts strongly with the electromagnetic field of the atomic nucleus and spontaneously annihilate into an electron-positron pair [9, 10]. In clinical radiotherapy photon energies of about 4-22 MV are used. In this range the Compton Effect is dominating, as can be interpreted from Figure 1.

In these processes free electrons are created, and it is these secondary particles that are responsible for most of the biological damage in photon therapy. Hence, photons are often referred to as indirectly ionizing particles.

A photon beam will decrease in intensity as it traverses the absorbing medium. The mathematical description of photon intensity, as a function of distance travelled in an absorbing medium, $I(x)$, is given as:

$$I(x) = I_0 e^{-\mu x} \quad (1)$$

Where I_0 is the photon intensity before the beam enters the medium, x is the distance traversed in the medium by the beam and μ is the linear attenuation coefficient of the target material [9, 10]. The linear attenuation coefficient is also called *macroscopic cross section*, which in turn reflects the probability of a certain process to occur [10]. Which of the processes will occur is highly dependent on the energy of the photon as seen in Figure 1. The density (ρ) and atomic number (Z) of the target material also affect the probability of the different processes.

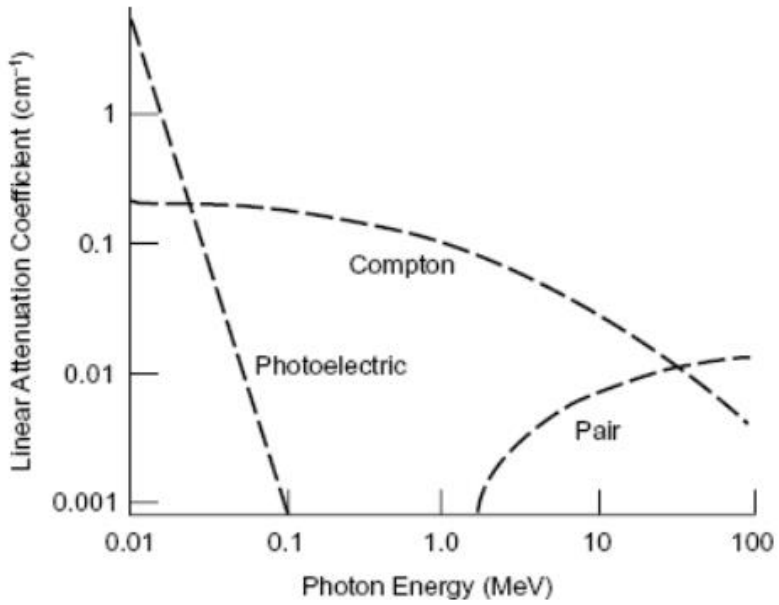

Figure 1: Linear attenuation coefficient versus photon energy for the relevant photon interaction processes [11].

This thesis uses photon radiation as a reference to the radiation of protons.

2.1.2 Proton interaction with matter

A proton is a heavy, positively charged particle. A heavy particle has a significant rest mass compared to the rest mass of an electron. Protons interact with matter primarily

through three mechanisms: Coulomb interactions with orbital electrons, Coulomb interactions with nuclei, and nuclear interactions [9, 12, 13].

In Coulomb interactions with orbital electrons, the proton hits an orbital electron and transfers a fraction of its energy to it. The electron can be excited to a higher shell or, if the energy transferred is high enough, the electron can escape the atom all together, creating an ion-pair. This process is the main contributor to the energy loss of the proton beam. The secondary electrons that escape their original atom are called *delta-rays* and will ionize and deposit energy a short distance from the protons' path. As the proton is far heavier than an electron, the proton's path will not be affected much from a single interaction [12].

If a proton travels close to a nucleus at a shallow angle it will be repulsed by the Coulomb force, since both the nucleus and the proton is positively charged. Through this interaction the direction of the proton beam will change slightly, however, the proton will not lose any energy [12]. Although this does not affect the proton beam significantly, if it occurs several times in the proton's path it can lead to non-negligible lateral spreading of the beam. This is called *"multiple Coulomb scattering"* [12].

Finally, the proton can interact directly with the nucleus if the angle of approach is narrow. A large portion of energy is then transferred to the nucleus from the incident proton, which can lead to scattering at a large angle. Nuclear scattering can be either elastic or non-elastic. In elastic nuclear scattering the nucleus will recoil, and the total kinetic energy will be preserved. In non-elastic nuclear scattering, the energy transferred from the proton to the nucleus can lead to secondary events such as disintegration of the nucleus into smaller fragments or the emission of prompt gamma rays. If the nucleus recoils, the proton will be absorbed at the point of interaction, which will also happen to heavier fragments of a reduced nucleus. The scattered protons and secondary neutrons can travel a relatively large distance from the point of interaction, creating a *"halo"* of low dose deposition [12].

An illustration depicting the processes for proton interactions with matter is shown in Figure 2.

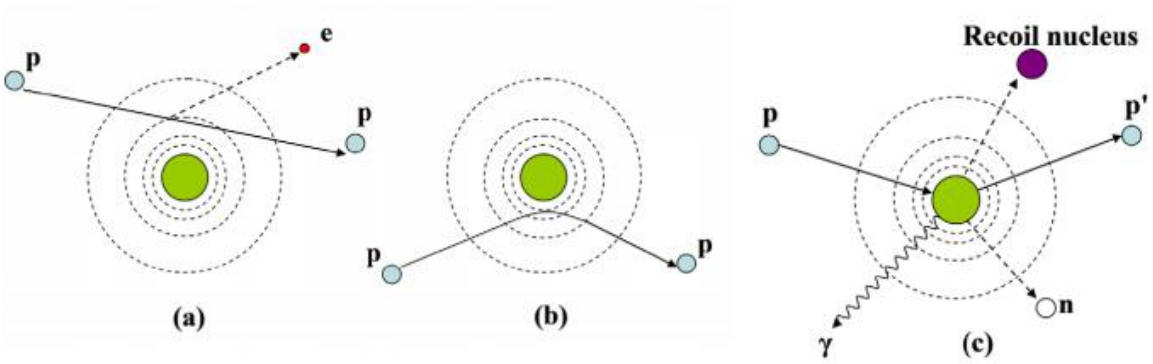

Figure 2: Illustrations of the main processes describing proton interactions with matter: (a) Coulomb interaction with orbital electrons, (b) Coulomb interaction with a nucleus, and (c) nuclear interaction [13].

2.1.3 Dosimetry

The measure of energy absorbed by an irradiated material is called absorbed dose. Dose is used for all types of ionizing radiation, all materials, and all energies, and it is the basic dose quantity used in radiotherapy. The absorbed dose D is given by the following equation:

$$D = \frac{d\epsilon}{dm} \tag{2}$$

Where $d\epsilon$ is the mean energy deposited to a material of mass dm. The unit of the absorbed dose is Gray (Gy) where Gray is equal to one Joule per kilogram $\left[\frac{J}{kg}\right]$.

2.1.4 Depth-Dose profile

Photon beams have a depth-dose curve that rises towards a maximum in the so called *dose build-up region* [9], before gradually decreasing, as can be seen in Figure 3. The

rise in dose is due to the production of free electrons through photoelectric effect, Compton scattering or pair production. The position of the dose maximum is dependent on the photon beam energy but is generally in radiotherapy located a few centimetres into the patient. The decrease in dose deposition, following the maximum, stem from the attenuation of the photon beam [9].

Protons, on the other hand, have a significantly longer build-up region. The dose deposited increases exponentially until it reaches a narrow maximum, the Bragg Peak (BP). Distal to the BP, the dose falls rapidly to zero. The shape is a result of the stopping power for protons, as explained by equation (3), which is inversely proportional to the velocity squared. In the shallow region, where the proton beam has the highest velocity, the deposited dose will be lowest. When the beam becomes attenuated, thus slowed down, the absorbed dose will increase until the velocity reaches zero and the protons are completely absorbed in the BP. The Depth-Dose profile is one of the main arguments for proton therapy over photon therapy. Heavier ions such as Carbon nuclei have a similar Depth-Dose curve to protons, but with a sharper BP. However, because heavier ions can be broken down into smaller fragments, a distal "tail" of dose is also deposited beyond the BP. This can lead to complications if the tumour is located close to an OAR.

By irradiating several times with different proton energies, dose can be deposited homogeneously to a larger area to form a Spread-Out Bragg Peak (SOBP). The Depth-Dose profiles of a photon beam, a monoenergetic proton beam and a SOBP proton beam are depicted in Figure 3.

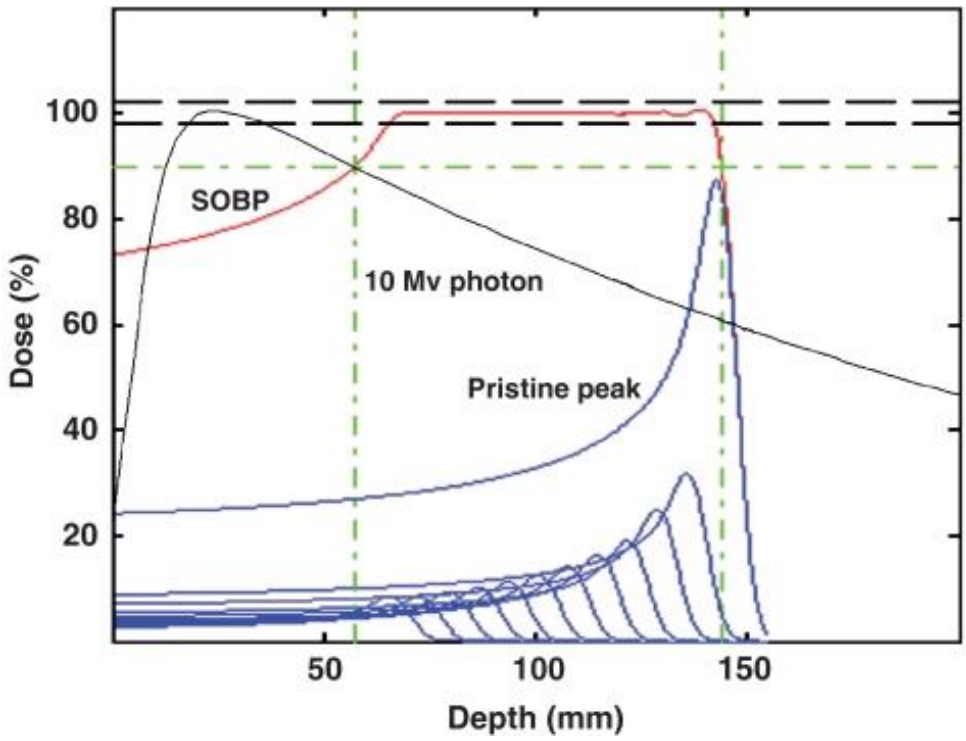

Figure 3: Depth dose profile of photons (black), monoenergetic protons (blue) and a Spread Out Bragg Peak (SOBP) (red) as a result of several monoenergetic irradiations [14].

2.1.5 Stopping power

As protons and other heavy charged ions travel through a medium they mainly lose energy due to collisions with orbital electrons (Chapter 2.1.2). Despite each individual collision contributing to a very small amount of energy lost, because of the high frequency of interactions per unit length, the loss of energy as a function of distance travelled becomes substantial. The energy lost by the particle to this effect per unit length traversed is known as *electric stopping power* and is described by the *Bethe-Bloch* formula [13, 15, 16]:

$$-\frac{dE}{dx} = 2\pi N_a r_e^2 m_e c^2 \rho \frac{Z}{A}\frac{z^2}{\beta^2}\left[\ln\left(\frac{2m_e\gamma v^2 W_{max}}{I^2}\right) - 2\beta^2 - \delta - 2\frac{C}{Z}\right] \qquad (3)$$

The variables in the *Bethe-Bloch* formula is described in Table 1.

The correction terms at the end of the *Bethe-Bloch* formula (δ and C) arise from relativistic theory and quantum mechanics and only become significant when very high or very low proton energies are used. From Equation (3) it can be seen that the stopping power is not dependent on the mass of the proton, but is proportional to the proton charge squared and inversely proportional to the proton velocity squared.

Table 1: Description, value, and unit of *Bethe-Bloch* variables.

Variable	Description	Value and unit
$\boldsymbol{dE}$	Change in energy	
$\boldsymbol{dx}$	Small part of particle track	
$\boldsymbol{N_a}$	Avogadro's constant	$6{,}022 * 10^{23} mol^{-1}$
$\boldsymbol{r_e}$	Electron radius	$2{,}818 fm$
$\boldsymbol{m_e}$	Electron mass	$0{,}511 MeV/c$
$\boldsymbol{c}$	Velocity of light in a vacuum	$2{,}998 * 10^{8} m/s$
$\boldsymbol{\rho}$	Mass density of absorbing material	g/cm^3
$\boldsymbol{Z}$	Atomic number of absorbing material	
$\boldsymbol{A}$	Atomic mass of absorbing material	g/mol
$\boldsymbol{z}$	Charge of projectile particle (proton)	C
$\boldsymbol{\beta = v/c}$	Relativistic velocity of incident particle (proton)	
$\boldsymbol{\gamma}$	Lorentz factor $= 1/\sqrt{1-\beta^2}$	
$\boldsymbol{v}$	Velocity of incident particle (proton)	m/s
$\boldsymbol{W_{max}}$	Maximum energy transfer from a single collision	
$\boldsymbol{I}$	Mean excitation potential absorbing material	eV
$\boldsymbol{\delta}$	Density corrections	
$\boldsymbol{C}$	Shell corrections	

2.1.6 Linear Energy Transfer

Linear Energy Transfer (LET) is used to describe the ionisation density in a particle track as a function of depth travelled in the tissue. LET is strongly depending on the energy of the particle, and its unit is $\frac{keV}{\mu m}$. Unrestricted LET (LET_∞) will be the same as the stopping power of the particle, restricted LET (LET_Δ), however, excludes the secondary electrons with energies over a threshold, Δ [17]. LET_Δ is given by the following equation [18]:

$$LET_\Delta = L_\Delta = \frac{dE_\Delta}{dx} \tag{4}$$

where dE_Δ is the average loss of energy by the emitted particles as it traverses the tissue and dx is the traversed distance.

To better reflect the biological effect of the radiation, the dose averaged LET (LET_d) is mostly used in radiation therapy today. LET_d is given by the following equation:

$$LET_d(z) = \frac{\int_0^\infty S_{el}(E)D(E,z)dE}{\int_0^\infty D(E,z)dE} \tag{5}$$

where $S_{el}(E)$ is the electronic stopping power of primary charged particles with kinetic energy E, and $D(E,z)$ is the absorbed dose contributed by primary charged particles with kinetic energy E at location z [19]. Clinically only LET_d values lower than 20 $\frac{keV}{\mu m}$ are used.

LET_d is proportional to the electronic stopping power, as shown by equation (5), which in turn is inversely proportional to the velocity of the beam. Therefore, the LET_d will drastically increase when the beam slows down at the distal edge the beam track, close to the BP. This is illustrated in Figure 4.

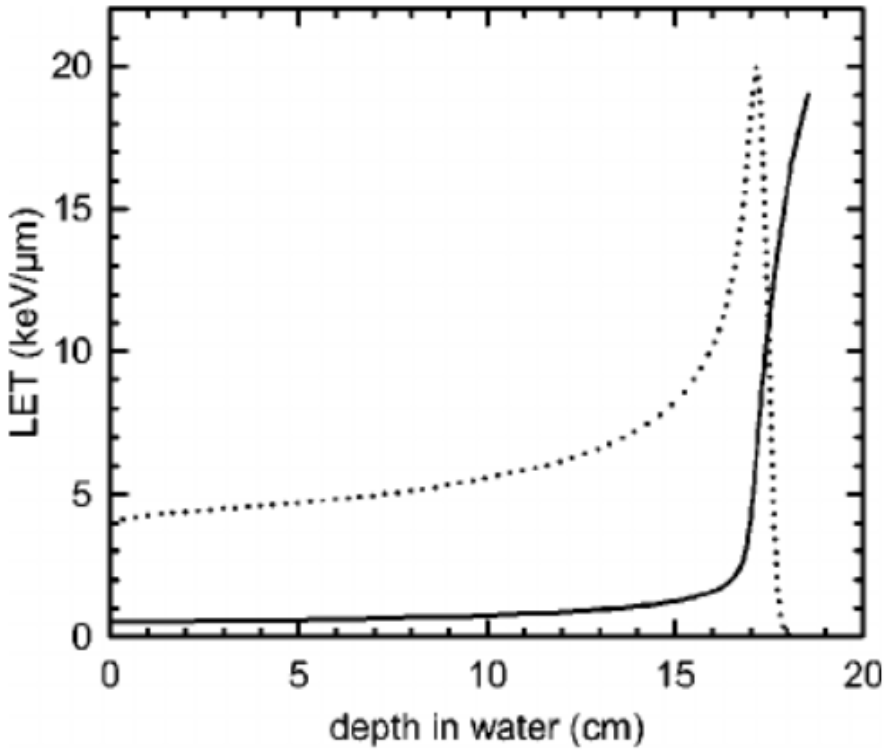

Figure 4: Dose averaged LET (solid) depth distribution of a monoenergetic proton beam with an initial energy of 160 MeV with corresponding dose (dotted) depth distribution resulting in a Bragg Peak [20].

2.2 Radiobiology

Radiobiology is the study of the effect of ionizing radiation on living organsims, where in this thesis, the effect on tumour and healthy tissue is of interest. In this chapter, some key concepts in radiobiology will be described.

2.2.1 Colonial cell experiments

The data that was used in this thesis was gathered from *in vitro* (Latin: in the glass) experiments, meaning that cell colonies are grown in petri-dishes, and subsequently irradiated. From these experiments, the Survival Fraction (SF) and radiation dose (D)

can be found. This process can be repeated for several different dose values. When several dose and SF pairs are obtained, they can be plotted in a graph and by regression one can create a cell survival curve [21]. By doing this for several types of radiation, radiation techniques, and cell lines, several cell survival curves can be created and compared, as shown in Figure 5. For *in vitro* experiments, several cell lines can be studied, where the most commonly used are Chinese hamster ovary (CHO) cells.

The alternative method is *in vivo* (Latin: within the living) experiments, meaning experiments on whole living organisms, such as animals, plants or even humans. Since *in vivo* experiments represent the actual conditions of living organisms, whereas *in vitro* only apply to the specific cells that are being irradiated, it is mostly agreed that *in vivo* experiments are better to represent clinical scenarios. However, as *in vitro* experiments are far easier to perform in a laboratory, and are less likely to harm sentient living organisms, *in vitro* is the method most commonly used today.

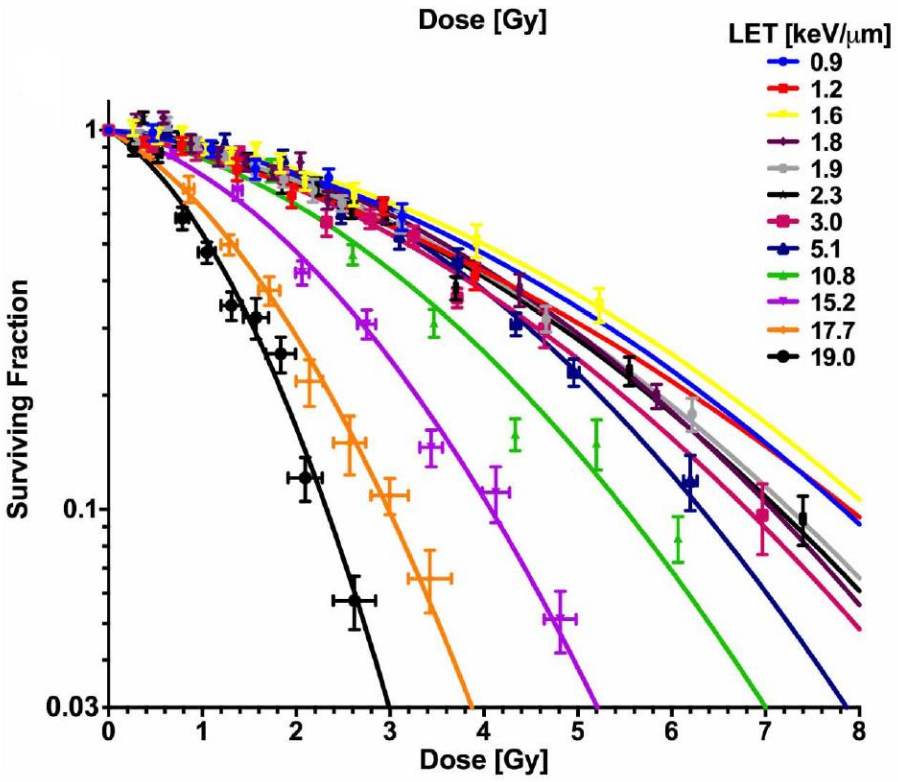

Figure 5: Example of cell survival curves derived from irradiation with several different doses and LET values [22].

2.2.2 The Linear-Quadratic model

The linear-Quadratic (LQ) model is a mathematical model used to describe the biological response of cells to radiation, by fitting the logarithm of the survival fractions as a function of dose from colonial cell experiments, as in Figure 5, to a second order polynomial [23]. The survival fraction is here given as:

$$-\ln(SF) = \alpha D + \beta D^2 \rightarrow SF = e^{(-\alpha D - \beta D^2)} \quad (6)$$

where α and β are tissue specific parameters. The initial slope, and linear component of the survival curve is determined by α. The linear component represents lethal cell damage in form of single-track events, in where a single particle from the beam is responsible for one complete DSB. The curvature and quadratic component of the survival curve, on the other hand, is determined by β. This represents repairable cell damage in form of double-track events, meaning two separate particles from the beam are responsible for a SSB each resulting in similar lethal damage to the DNA helix [23].

The ratio of these tissue specific parameters, $\frac{\alpha}{\beta}$, is defined as the dose where the linear and quadratic components contribute to the same amount of cell killings. Early responding tissue, such as skin and most tumours, has a relatively high $\frac{\alpha}{\beta}$ ratio in the range 7-20 Gy. Early responding tissue usually expresses damage within days or weeks. Late responding tissue has a low $\frac{\alpha}{\beta}$ ratio of 0.5 - 6 Gy. Late responding tissue expresses damage after several months or even years [23]. An illustration depicting the interpretation of the linear and quadratic components of cell survival, as well as the $\frac{\alpha}{\beta}$, is shown in Figure 6.

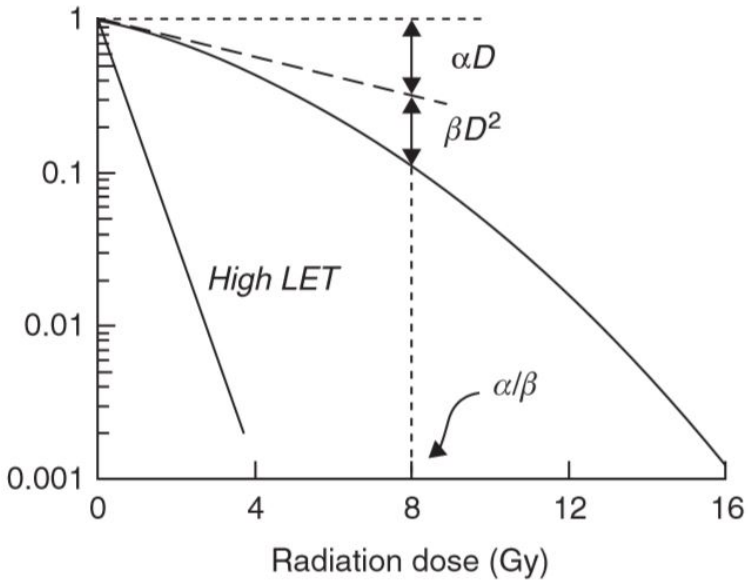

Figure 6: Example of cell survival curves with interpretation of the linear and quadratic components of the LQ model as well as their ratio. The straight line depicts a typical survival curve for high LET radiation [23].

Data suggests that the LQ model fits accurately in the middle dose ranges, however, for doses lower than 1-2 Gy [21, 23, 24] and higher than 5-6 Gy [21, 23], some modifications to the model could provide a better fit. In the low dose range, the induced repair (IndRep) model has been proposed, where both the linear and quadratic components of the LQ-model are corrected based on the dose where the cells stop being hypersensitive and start becoming more radioresistant [21, 23]. In the high dose region, the Linear-Quadradic-Cubic (LQC) model has been proposed. In this model a cubic component is added to the LQ-model to correct for the effect of cell overkill on the survival fraction [23].

In this thesis, only the original LQ-model is regarded.

2.2.3 Relative Biological Effectiveness

When comparing different radiation modalities, such as photons, protons, and heavier ions, and specifically the difference in their ability to kill cells, the Relative Biological Effectiveness (RBE) is used. Clinically a constant RBE of 1.1 is used in treatment planning for proton therapy, but there is general agreement that the RBE is not constant

and often higher than 1.1, especially for higher LET values. The general way to calculate RBE is by comparing the dosage necessary to kill a certain percentage of the cells for two different radiation modalities, thus referring to a specific survival fraction, e.g. $SF = 10\%$. These dose values can be gathered from cell survival curves (Figure 5) containing data from the modalities one wants to compare. When comparing protons to a reference radiation, e.g. Cobolt-60 γ-rays, the equation for RBE based on the endpoint survival is given by:

$$RBE(D, SF = 10\%) = \frac{D_{^{60}Co}(SF = 10\%)}{D_p(SF = 10\%)} \tag{7}$$

where $D_{^{60}Co}(SF = 10\%)$ is the dosage needed by the reference radiation Cobolt-60 to kill 90% of the cells and $D_p(SF = 10\%)$ is the dosage needed by the proton radiation to kill 90% of the cells. RBE is, as the name indicates, a ratio and is therefore unitless.

The RBE depends on a number of variables such as absorbed dose, LET, type of radiation, radiation quality, e.g. LET, type of cells and endpoint [25-27]. The RBE can also be calculated using parameters from the LQ model by the following equation [25, 28]:

$$RBE\left(D_p, \alpha, \beta, \alpha_x, \beta_x\right) = \frac{1}{2D_p}\left(\sqrt{\left(\frac{\alpha_x}{\beta_x}\right)^2 + 4D_p\frac{\alpha_x}{\beta_x}\frac{\alpha}{\alpha_x} + 4D_p^2\frac{\beta}{\beta_x}} - \frac{\alpha_x}{\beta_x}\right) \tag{8}$$

where D_p is the physical dose deposited by the protons, α and β are the tissue specific LQ parameters for the proton and, α_x and β_x are the tissue specific LQ parameters for the reference radiation. As the RBE is known to be inversely proportional to the dose, the maximum and minimum values of RBE is defined to be in the low and high dose limits respectively and is given by:

$$RBE_{max} = \lim_{D\to 0} RBE = \frac{\alpha}{\alpha_x} \tag{9}$$

and:

$$RBE_{min} = \lim_{D \to \infty} RBE = \sqrt{\frac{\beta}{\beta_x}} \tag{10}$$

This can be seen from the LQ-model, where the linear component is clearly dominant in the low dose region, while the quadratic component dominates the high dose region.

By inserting equations (9) and (10) into equation (8), the RBE can be written as [29, 30]:

$$RBE\left(D_p, \left(\frac{\alpha}{\beta}\right)_x, RBE_{max}, RBE_{min}\right) = \frac{1}{2D_p}\left(\sqrt{\left(\frac{\alpha}{\beta}\right)_x^2 + 4D_p\left(\frac{\alpha}{\beta}\right)_x RBE_{max} + 4D_p^2 RBE_{min}^2} - \left(\frac{\alpha}{\beta}\right)_x\right) \tag{11}$$

This equation is common for all LQ-based RBE models, however, they differ in their definition of RBE_{max} and RBE_{min} [25].

To account for the difference in killing efficiency between radiation modalities, treatment plans are usually not based on the physical dose, but rather the RBE-weighted dose (D_{RBE}), also called biological dose:

$$D_{RBE} = RBE \times D \tag{12}$$

where D is the physical (absorbed) dose. The units for biological dose are also Gray (Gy), however, to distinguish it from the physical dose it is usually written as $Gy(RBE)$.

LQ based RBE models are generally derived by calculating the RBE_{min}, and RBE_{max}, using equations (9) and (10), then fitting these values to custom equations using a regression algorithm. The equation used in this process can vary from model to model, however, for RBE_{max} the equation is usually dependent on LET_d and $\left(\frac{\alpha}{\beta}\right)_x$, whereas the RBE_{max} is usually shown to be increasing with increasing LET_d, with a steeper

slope for low $\left(\frac{\alpha}{\beta}\right)_x$ values. The RBE_{min} is often assumed to be equal to one [31]. As different databases are used to derive these models, and the definitions of RBE_{min}, and RBE_{max} differ, the dose distributions recalculated using these models also differ, as depicted in Figure 7.

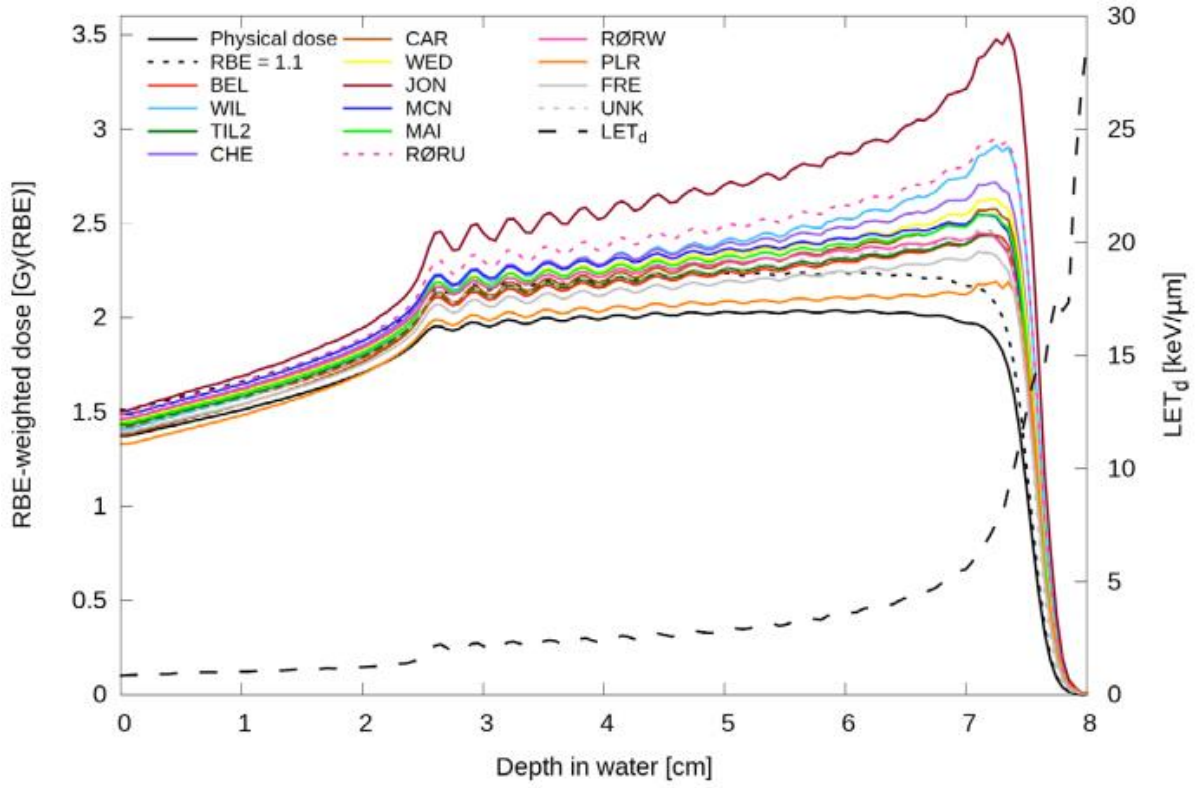

Figure 7: Depth dose distribution of a SOBP in a water phantom recalculated with various RBE models. LET_d is shown by the dotted black line. An $\left(\frac{\alpha}{\beta}\right)_x$ of 3.76 Gy was used in the calculations [31].

2.3 Cluster analysis

A cluster analysis is a tool used to find groups in sets of data [32]. These groups, or clusters, are determined only by the information found in the dataset, and the relationships within the data. The goal of a cluster analysis is to separate the data into

clusters, where the data within a certain cluster is similar (or related) and the data in one cluster is different (or unrelated) compared to data in a different cluster [33].

Cluster analysis has a wide range of uses. It is used in several fields, such as Biology, Business, Climate research, Medicine and Psychology. One of the most famous examples of clustering is the Taxonomy of all living things. This is a hierarchical classification where each individual species is considered as an original singleton cluster. Species are merged into genus based on their shared characteristics. Genus are, consecutively, merged into family, then order and after several iterations everything is categorized as life in the final cluster [33, 34].

Some popular clustering algorithms are K-means clustering, Agglomerative Hierarchical clustering and Density-based clustering of applications with noise (DBSCAN). K-means clustering demands a user-specified number of clusters (K). K initial centroids are then chosen and all datapoints are assigned to the closest centroid. New centroids are then computed, and the data is reassigned to its current closest centroid. This process is repeated until the centroids remain unchanged between iterations. Agglomerative Hierarchical clustering considers every single datapoint as a singleton cluster. The two closest clusters are then merged into one and further considered as one entry. This process can be repeated as many times as necessary, even until every datapoint in the database is merged into a single cluster. DBSCAN is a density-based algorithm. Regions in the dataset with a high density of datapoints, separated by low-density regions, are defined as a cluster. Low-density regions are, however, considered noise and can be omitted by the algorithm [33].

Cluster analysis has also been utilized in radiotherapy, for example to classify radioresistant and sensitive cell lines [35], segmenting tumour subregions (Supervoxels) and grouping these subregions into phenotypic, meaning observably distinguishable, clusters [36], and to establish clusters based on combinations of treatment options [37].

3. Material and Methods

3.1 The database

A database containing data from 98 *in vitro* cell survival experiments was collected for this project, including a total of 730 discrete data points from proton irradiation. What is referred to as an experiment is here one row in the final database, represented by one cell survival curve with all coresponding relevant data. The database is based on the database used in by *Rørvik et al 2019* [8], combined with data from *Paganetti et al 2014* [26] and finally from the Particle Irradiation Data Ensemble (PIDE 3.1) [38] database. Firstly, only experiments using proton radiation were included, additionally, all non-monoenergetic experiments were excluded. The database was then restricted to only contain experiments with LET_d values lower or equal to $20\,\frac{keV}{\mu m}$. An important difference from Rørvik's database, compared to the one used in this thesis, is that Rørvik restricted the database to contain only experiments with $\left(\frac{\alpha}{\beta}\right)_x$ values lower or equal to $5\ Gy$. This was because Rørvik's database was representing late responding tissue, but for this study that was not the case, so this restriction was removed.

The database includes information on the type of reference radiation used in the experiments, as well as α- and β-values for both the reference radiation and the proton radiation. The LET_d values were provided in both Rørvik's and Paganetti's databases and is, thus, also included here. The $\left(\frac{\alpha}{\beta}\right)_x$, RBE_{min} and RBE_{max} values were calculated using equations (9) and (10). Finally, the dose values used to irradiate the cells and the corresponding cell survival fractions were retrieved from Rørvik and PIDE databases. For experiments not included in Rørvik's or PIDE's databases, WebPlotDigitizer Version 4.2 [39] was used to extract cell survival for different dose levels from curves in the relevant papers.

The data in our database consists of 98 cell survival curves from the following 27 papers: *Baggio et al 2002* [40], *Belli et al 1998* [41], *Belli et al 2000* [42], *Bettega et al 1998* [43], *Bird et al 1980* [44], *Chaudhary et al 2014* [45], *Folkard et al 1996* [46], *Folkard et al 1998* [47], *Fuhrman Conti et al 1988* [48], *Green et al 2001* [49], *Green at al 2002* [50], *Guan et al 2015* [22], *Hei et al 1988a*[51], *Hei et al 1988b* [52], *Howard et al 2017* [53], *Jaynes et al 2013* [54], *Moertel et al 2004* [55], *Patel et al 2017* [56], *Perris et al 1986* [57], *Petrovic et al 2006* [58], *Prise et al 1990* [59], *Schettino et al 2001* [60], *Schuff et al 2002* [61], *Sgura et al 2000* [62], *Slonina et al 2004* [63], *Wainson et al 1974* [64], and *Wera et al 2013* [65].

The database includes experiments using the following photon reference radiations: ^{60}Co, ^{137}Cs, 6 MV photons, 120 kVp x-rays, 200 kVp x-rays, 225 kVp x-rays, 240 kVp x-rays, 250 kVp x-rays and 300 kVp x-rays. As correcting RBE directly from the reference radiation is not feasible [26], the LET_d values were normalized to Cobalt-60 instead using the following equation [25, 66]:

$$LET^* = LET_p - LET_x + LET_{^{60}Co} \quad (13)$$

Where LET^* is the normalized LET_d value, LET_p is the LET_d value of the proton beam in each experiment, LET_x is the LET_d value of the reference radiation used in the corresponding experiment and $LET_{^{60}Co}$ is the LET_d value of ^{60}Co. The LET_x values used for normalisation are listed in Appendix A.

3.2 Cluster analysis

We wished to investigate the effect the ranges of the input data have on the RBE modelling process. To achieve this, our database was divided into smaller fractions with similar experimental data within each fraction [67].

A cluster analysis was performed on the data set using the Two-Step clustering algorithm in the IBM SPSS® Statistics software version 25. Two-Step clustering has

as the name indicates two steps, a non-hierarchical and a hierarchical step. First, in the non-hierarchical part, or sequential algorithm, the algorithm runs through all the experiments in the database in order, it then adds the current experiment to an existing pre-cluster if the distance measure chosen for the algorithm is within a threshold distance determined by the algorithm. If it is not, the algorithm creates a new pre-cluster and adds the experiment to it. Since this step depends on the order the input is read, the data was randomized prior to running the algorithm. The pre-clusters are considered the input to the hierarchical part of the algorithm, and cannot be split into different clusters during this step [68].

Early in the project, it was not yet decided if categorical variables, e.g. the type of reference radiation, would be included as an input variable. A distance measure of Log-Likelihood distance was, therefore, chosen for this cluster analysis, because it could be used for both continuous, and categorical variables, in case of changes in the choice of input variables. The Log-Likelihood distance measure between two clusters is related to the decrease in Log-Likelihood as they are merged into one cluster. The distance between clusters a and b, $d(a, b)$, are calculated in SPSS by the following equation:

$$d(a,b) = \xi_a + \xi_b - \xi_{<a,b>} \tag{14}$$

Where:

$$\xi_v = -N_v \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 - \hat{\sigma}_{vk}^2) + \sum_{k=1}^{K^B} E_{xk} \right) \tag{15}$$

Here N_v is the number of data records in cluster v, $\hat{\sigma}_k^2$ is the estimated variance of the *k*th variable across the entire dataset, $\hat{\sigma}_{vk}^2$ is the estimated variance of the *k*th variable in cluster v. K^A and K^B represent the total number of continuous and categorical variables, respectively. As we did not use any categorical variables in our analysis the summation to the right in equation (15) can be neglected. The notation $< a, b >$ describes a merged cluster consisting of the previously separate clusters a and b [68].

The variables used to create the clusters were the α- and β-values from the proton radiation, the normalized LET_d-value and the lowest dose from each cell line in the data set, D_{min}. The $\left(\frac{\alpha}{\beta}\right)$ from the reference radiation as well as the RBE_{max} and RBE_{min} values, were used as additional evaluation fields during the analysis in order to spot trends within these variables based on the cluster distributions. An evaluation field is a variable which is not used to determine the cluster memberships, however considered when comparing the clusters. The $\left(\frac{\alpha}{\beta}\right)_x$ was not included as an input variable, because the same reference radiation was used in several experiments, and therefore, if it would be used to determine the cluster memberships, the clusters could be determined by these values alone, meaning the $\left(\frac{\alpha}{\beta}\right)_x$ would be a so called swamping variable. A swamping variable is a variable that essentially hijacks the clustering when it comes to cluster membership partitioning. The RBE_{max} and RBE_{min} values were not included as input variables, because we wanted to find trends in these variables based on the cell survival data included in the database.

To determine the ideal amount of clusters, a combination of Akaike Information Criterion (AIC) [69] and a silhouette score was used. AIC is an estimation of information lost in a model, where a low AIC value translates to little information lost, and a high AIC value indicates a high amount of lost information. The AIC value for a model with J clusters is calculated by SPSS using the following equation:

$$AIC(J) = -2l_J + 2m_J \tag{16}$$

Where m_J is the number of independent parameters, and:

$$l_J = \sum_{v=1}^{J} \xi_v \tag{17}$$

where again v and ξ_v are from Equation (15) [69-71], and:

$$m_J = J\left(2K^A + \sum_{k=1}^{K^B}(L_K - 1)\right) \quad (18)$$

An alternative to AIC was the Schwarz's Bayesian information criterion (BIC). This is calculated by SPSS with the following equation:

$$BIC(J) = -2l_J + m_J \log(N) \quad (19)$$

Where N is the total number of experiments in the analysis, l_j is given by equation (17), and m_j is given by equation (18).

Similar to Equation (15) the righthand sum within the parenthesis can be neglected as no categorical variables were used [68]. It is believed that BIC is slightly inferior to AIC [70, 72, 73], therefore AIC was most heavily considered when deciding the preferable number of clusters in the analysis. BIC was included for comparison.

The silhouette score describes how similar the data within a cluster is, compared to all other clusters. The silhouette score (s) is calculated in SPSS by the following equation:

$$s = \frac{B - A}{\max(A, B)} \quad (20)$$

where A is the average distance between a given datapoint and the centroid of the cluster the datapoint is in, and B is the smallest average distance between the same datapoint to the centroid of other clusters [33, 68, 74, 75]. A silhouette score ranges between -1 and 1, where 1 translates to a perfect model, however, an average silhouette score of higher than 0.5 is regarded as a "good" model and a score between 0 and 0.5 is considered a "fair" model. A model is considered as "bad" if its silhouette score is lower than zero [32, 68]. The average Silhouette score of each individual datapoint is

what is considered in the final model. Silhouette scores are only applicable for models with two or more clusters.

When deciding the ideal number of clusters, the Predictor Importance was also taken into consideration. Predictor Importance is a measure of the ability of a variable to differentiate clusters. A high Predictor Importance value indicates that the variation in the variable is caused by an underlying difference rather than mere chance. The Predictor Importance value (VI_i) is normalized so that the highest value is one, and calculated by SPSS using the following equation:

$$VI_i = \frac{-\log(sig_i)}{\max_{j\in\Omega}\left(-\log\left(sig_j\right)\right)} \tag{21}$$

where Ω is the set of variables and evaluation fields, sig_i is the significance or p-value of variable i, determined by applying the F-test for continuous fields and Pearson's chi-square for categorical fields, such as reference radiation [32, 68]. In contrast to the silhouette score, the predictor importance does not have numerical definitions where it is defined as "good" or "bad". However, a model where the predictor importance of each variable is close to equal and close to one is preferred.

AIC values for 1 to 15 clusters, and Silhouette scores for 2 to 15 clusters were calculated and plotted to select the ideal number of clusters. BIC values were also calculated and plotted for comparison. For the number of clusters deemed ideal given the evaluation parameters, the clusters were plotted and compared. In addition to the input variables, the resulting clusters were also compared in regard to the evaluation fields, $\left(\frac{\alpha}{\beta}\right)_x$, RBE_{max} and RBE_{min}.

Based on the results of this analysis, a secondary cluster analysis was performed where four extreme values were excluded, because they do not represent clinically relevant cells and dose deliveries. One experiment from *Petrovic et al* [58] was excluded for its unusually high D_{min} value of $8.015\ Gy$. Another Cell survival curve from *Baggio et al*

[40] was excluded where the $\left(\frac{\alpha}{\beta}\right)_x$ value was $69.5\ Gy$, which is significantly higher than typical clinical values. The final experiments excluded were two cell survival curves from *Schuff et al* [61], because of their extremely low reference α value of $0.004\ Gy$, resulting in an unusually low $\left(\frac{\alpha}{\beta}\right)_x$ value of $0.051\ Gy$ and RBE_{max} values of 35 and 80, respectively. Thus, the final database ended up comprised of 94 experiments.

3.3 Model comparison

Eleven existing RBE models were compared to the clusters from the analysis performed in this study. The following models were compared: *Belli et al 1997* [76], *Wilkens and Oelfke et al 2004* [28], *Tilly et al 2005* [67], *Chen and Ahmad et al 2012* [77], *Carabe et al 2012* [78], *Wedenberg et al 2013* [79], *Jones et al 2015* [80], *McNamara et al 2015* [81], *Mairani et al 2017* [82], *Rørvik et al 2017* [25] and *Peeler et al 2016* [83].

The comparison was done to see if any of the clusters would be over- or underrepresented in any of the models, and to find out if any trends in the models are in agreement with the cluster model.

3.4 RBE modelling/estimation

To further investigate trends in the database, the curve fitting tool in MATLAB version R2018a (MathWorks, Natick, MA, USA) [84] *(cftool())* was utilized to fit the datapoints to the RBE_{max} and RBE_{min} model equations:

$$RBE_{min} = 1 + c_{min} \times LET^{*} \tag{22}$$

And:

$$RBE_{max} = 1 + \frac{c_{max}}{\left(\frac{\alpha}{\beta}\right)_x} \times LET^* \tag{23}$$

Where c_{min} and c_{max} are the fitting variables of the resulting models. The custom equation option in *cftool()* was utilized to fit the data to the relevant equation.

The RBE_{min}, as mentioned earlier, is often assumed to be equal to one. However, to investigate if this is a valid simplification, we decided to fit it to an equation linearly dependent on the LET_d. If the assumption is valid, we would expect fitting variables close to zero. The trends for RBE_{max} are generally, as mentioned in Chapter 2.2.3, increasing with LET_d, where the slope gets steeper for low $\left(\frac{\alpha}{\beta}\right)_x$ values. The general form of equation (23) is the most frequently used equation to fit RBE_{max} to , however, some models choose to let the regression algorithm decide the constant term, instead of setting it to one. Relevant model equations can be found in Appendix C.

Both the RBE_{min} and RBE_{max} were fitted to datapoints in each of the five clusters. A final curve fit using all datapoints was also performed. This model is throughout the rest of the thesis referred to as the "All clusters" model.

To represent "Goodness of fit" the R-squared values were utilized. This statistic represents the relative fraction of variation about the average in the data, that can be explained by the model. An R-squared value is normally between one and zero, where a value of one means that all variation can be explained by the model. The 95% confidence interval was included as a measure of margin of error.

3.5 FLUKA simulation and dose calculations

We further wish to investigate the dose and RBE distributions resulting from the different clusters found in this thesis to see if the data included in each model/cluster led to differences in dose, RBE and hence biological damage.

An existing proton treatment plan created in the Eclipse treatment planning system (Varian Medical Systems, Palo Alto, CA, USA) was provided. This plan was designed to deliver a dose of $2\ Gy(RBE)$ to a planning target volume (PTV) within a water phantom, assuming an RBE of 1.1 as is used in current clinical practice. The treatment plan information was given in Digital Imaging and Communications in Medicine (DICOM) format and included information such as CT images of the water phantom, a $4\ cm \times 4\ cm \times 4\ cm$ PTV, and treatment field specific information such as gantry angle and beam energy. Using an in-house Python script [85], the treatment plan information was converted into a format readable for the FLUKA Monte Carlo (MC) code [86, 87] called an input file. FLUKA also has a graphical user interface called FLAIR (FLUKA Advanced Interface) [88] which can be used to further set up the simulation environment. As FLUKA cannot read the DICOM CT images directly, FLAIR was used to convert the images into a so-called voxel file. This file contains material and density information for each voxel and was used during the simulations.

An executable file was created using the source and fluscw FLUKA user routines. The source routine reads specific information from the treatment field such as spot positions, spot sizes and beam divergence. As FLUKA and the DICOM format employs different coordinate systems, the information was converted into FLUKA format by the source routine. The fluscw routine was also used to obtain physical parameters such as the LET as well as to calculate the dose given by $fluence \times LET/\rho$, where ρ is the density of the material. We used a homogenous water phantom, so ρ is here the density of water, $1\ g/cm^3$ [85, 89]. The irradiation of the water phantom was simulated using 10^6 primary particles, creating scoring files for dose and LET in ASCII format. For all simulations, FLUKA version 2011.2x.8 and FLAIR version 2.3-0 was used.

Furthermore, the ASCII files created by the FLUKA simulation were converted back into DICOM files to make them compatible with the scripts used to create plots. In this process the biological dose for $RBE = 1.1$, LET_d, and the unweighted Rørvik [25]

model were exported along with the cluster models created in this thesis. The unweighted Rørvik model was chosen because of its high amount of common datapoints with our database. This conversion process essentially recalculated the dose within each voxel using equation (11) where D_p is the dose calculated based on the fluence, and RBE_{min} and RBE_{max} originated from the relevant models. The $RBE = 1.1$ model assume a $RBE_{min} = RBE_{max} = 1.1$. The model equations for the Rørvik unweighted model can be found in Appendix C, and the model equations for the cluster models can be found in Chapter 4.4. The calculations were done for $\left(\frac{\alpha}{\beta}\right)_x = 2\ Gy$, representing late responding tissues, and the LET_d values used to calculate RBE_{min} and RBE_{max} within each voxel were directly obtained from the FLUKA simulations. These new recalculated DICOM files were subsequently used to plot the dose and RBE distributions along with the LET_d distribution.

Cumulative Dose Volume Histograms (DVHs) were then created based on the recalculated DICOM files. In a cumulative DVH, the height of each bin along the x-axis represent the relative volume of the chosen structure that receives an equal or greater dose [9].

Throughout this process, existing in-house scripts from our group, with some project specific modifications, were used. The scripts were run in Python versions 2.7 and 3.6. A short description of the scripts used throughout this process can be found in Appendix D.

4. Results

4.1 Database calculations

From the calculations mentioned in Chapter 3.1 the database included $\left(\frac{\alpha}{\beta}\right)_x$ values in the range of $0.05 - 69.5\ Gy$, however, as further datapoints were excluded, this range was restricted to $1.2 - 28.5\ Gy$. Normalized LET^* values ended up in the range $0.4 - 19.8\ \frac{keV}{\mu m}$, both before and after the new restrictions. The calculated RBE_{min} range of $0 - 3.4$ was not affected by the exclusion either. The RBE_{max} values were originally in the range $0 - 80$, but following the restrictions a final RBE_{max} range of $0 - 7.2$ was obtained, due to the now excluded experiments.

38 of the total 98 experiments in the database have an RBE_{min} value higher than its RBE_{max} value, where one was the excluded *Baggio et al 2002* [40] experiment, resulting in 37 out of the 94 experiments in the restricted database having $RBE_{min} > RBE_{max}$. These experiments included all 12 experiments from *Patel et al 2017* [56], 13 experiments from *Guan et al 2015* [22], two from *Bird et al 1980* [44], two from *Green et al 2002* [50] and one experiment each from *Fuhrman Conti et al 1988* [48], *Wainson et al 1972* [64], *Sgura et al 2000* [62], *Jeynes et al 2013* [54], *Slonina et al 2014* [63], *Perris et al 1986* [57], *Belli et al 2000* [42] and *Folkard et al 1989* [47].

The calculated LET^*, RBE_{min} and RBE_{max} values for all experiments can be found in Table 2.

Table 2: The data used in this thesis listed in the order of ascending cluster membership. Within each cluster membership the data is listed by ascending $\left(\frac{\alpha}{\beta}\right)_x$.

Reference Radiation	$\left(\frac{\alpha}{\beta}\right)_x$ $[Gy]$	*LET** $\left[\frac{keV}{\mu m}\right]$	RBE_{min}	RBE_{max}	*Reference*	*Cluster membership*
60-Co	1.790	0.420	1.072	2.925	*Green et al 2002*	1
225kVp	1.833	0.392	1.033	1.273	*Chaudhary et al 2014*	1
225kVp	1.833	3.302	1.041	1.545	*Chaudhary et al 2014*	1
225kVp	1.833	6.282	1.088	2.000	*Chaudhary et al 2014*	1
225kVp	1.833	11.182	0.866	4.000	*Chaudhary et al 2014*	1
225kVp	1,833	17.282	0.365	7.000	*Chaudhary et al 2014*	1
60-Co	1.889	0.420	0.923	2.373	*Green et al 2001*	1
240kVp	2.708	9.408	0.901	2.462	*Folkard et al 1996*	1
137-Cs	3.892	0.390	0.658	0.757	*Wainson et al 1972*	1
60-Co	3.957	0.420	0.824	1.111	*Green et al 2002*	1
250kVp	4.040	8.555	2.332	1.975	*Bird et al 1980*	1
6 MV	4.081	1.980	0.959	1.464	*Howard et al 2017*	1
6 MV	4.081	3.560	1.151	1.609	*Howard et al 2017*	1
6 MV	4081	4.550	1.174	2.245	*Howard et al 2017*	1
6 MV	4.081	7.540	1.078	3.106	*Howard et al 2017*	1
6 MV	4.081	1.190	0.959	1.470	*Howard et al 2017*	1
6 MV	4.081	2.460	0.986	1.517	*Howard et al 2017*	1
6 MV	4.081	4.390	1.115	1.914	*Howard et al 2017*	1
6 MV	4.081	7.490	1.315	2.026	*Howard et al 2017*	1
250kVp	5.000	8.555	3.152	1.493	*Bird et al 1980*	1
60-Co	5.105	0.420	1.024	0.839	*Green et al 2002*	1
137-Cs	5.156	0.390	0.919	0.810	*Wainson et al 1972*	1
60-Co	5.405	0.420	1.000	0.890	*Green et al 2002*	1
120kVp	5.813	5.243	1.132	1.188	*Moertel et al 2004*	1
120kVp	6.387	5.243	1.231	2.545	*Moertel et al 2004*	1
250kVp	11.333	16.325	1.528	3.412	*Prise et al 1990*	1
6MV	12.379	8.100	0.983	1.134	*Slonina et al 2014*	1
6MV	13.063	8.100	1.137	1.005	*Slonina et al 2014*	1
6MV	16.805	8.100	0.988	1.241	*Slonina et al 2014*	1
60-Co	23.333	5.800	1.958	1.500	*Perris et al 1986*	1
137-Cs	2.286	9.600	0.474	6.196	*Hei et al 1988a*	2
225kVp	8.710	6.282	0.813	2.389	*Chaudhary et al 2014*	2
225kVp	8.710	11.182	1.129	3.148	*Chaudhary et al 2014*	2

Reference Radiation	$\left(\frac{\alpha}{\beta}\right)_x$ $[Gy]$	*LET** $\left[\frac{keV}{\mu m}\right]$	RBE_{min}	RBE_{max}	*Reference*	*Cluster membership*
225kVp	8.710	17.282	1.092	3.463	*Chaudhary et al 2014*	2
225kVp	8.710	3.302	0.992	1.889	*Chaudhary et al 2014*	2
225kVp	8.710	0.392	1.385	1.389	*Chaudhary et al 2014*	2
60-Co	18.387	19.700	0.000	1.526	*Belli et al 2000*	2
250 kVp	18.444	9.325	0.000	2.482	*Wera et al 2013*	2
137-Cs	1.220	17.300	1.906	6.560	*Guan et al 2015*	3
137-Cs	1.220	14.800	1.522	3.600	*Guan et al 2015*	3
137-Cs	1.220	10.400	1.148	2.380	*Guan et al 2015*	3
240 kVp	2.708	10.208	0.979	1.854	*Schettino et al 2001*	3
240kVp	2.708	17.108	0.764	3.462	*Folkard et al 1996*	3
200 kVp	2.804	19.236	0.967	3.636	*Belli et al 1998*	3
200 kVp	2.804	10.236	0.885	2.884	*Belli et al 1998*	3
200 kVp	2.804	6.936	0.722	2.240	*Belli et al 1998*	3
60Co	3.083	7.700	0.000	1.270	*baggio et al 2002*	3
137-Cs	3.494	10.400	1.362	1.097	*Guan et al 2015*	3
137-Cs	3.514	15.200	1.162	2.462	*Schuff et al 2002*	3
250kVp	4.074	16.325	1.291	3.182	*Prise et al 1990*	3
250kVp	4.074	16.325	1.700	1.182	*Folkard et al 1989*	3
250 kVp	4.390	9.325	0.855	1.722	*Hei et al 1988b*	3
300kVp	5.217	16.482	1.707	1.000	*Jeynes et al 2013*	3
60-Co	7.647	19.800	0.485	1.769	*Belli et al 2000*	3
60-Co	7.647	7.700	0.804	1.154	*Belli et al 2000*	3
60-Co	7.667	7.700	0.876	0.843	*Sgura et al 2000*	3
60-Co	15.000	19.700	1.541	1.792	*Bettega et al 1989*	3
60-Co	15.000	11.000	1.090	1.958	*Bettega et al 1989*	3
60-Co	18.387	7.700	1.723	0.719	*Belli et al 2000*	3
60-Co	23.333	12.100	0.000	3.071	*Perris et al 1986*	3
137-Cs	1.220	0.800	0.698	2.720	*Guan et al 2015*	4
137-Cs	1.220	0.500	0.826	1.540	*Guan et al 2015*	4
137-Cs	1.220	1.500	0.870	1.880	*Guan et al 2015*	4
137-Cs	1.220	2.600	0.897	2.220	*Guan et al 2015*	4
137-Cs	1.220	1.400	0.963	1.180	*Guan et al 2015*	4
137-Cs	1.220	1.900	0.883	1.920	*Guan et al 2015*	4
137-Cs	1.220	4.700	1.126	0.680	*Guan et al 2015*	4
137-Cs	1.220	1.200	0.812	1.340	*Guan et al 2015*	4
137-Cs	3.494	4.700	1.384	0.403	*Guan et al 2015*	4
137-Cs	3.494	2.600	1.227	0.710	*Guan et al 2015*	4
137-Cs	3.494	1.500	1.271	0.572	*Guan et al 2015*	4

Reference Radiation	$\left(\frac{\alpha}{\beta}\right)_x$ $[Gy]$	*LET** $\left[\frac{keV}{\mu m}\right]$	RBE_{min}	RBE_{max}	*Reference*	*Cluster membership*
137-Cs	3.494	1.400	1.271	0.517	*Guan et al 2015*	4
137-Cs	3.494	1.200	1.271	0.521	*Guan et al 2015*	4
137-Cs	3.494	0.500	1.081	0.924	*Guan et al 2015*	4
137-Cs	3.494	1.900	1.326	0.472	*Guan et al 2015*	4
137-Cs	3.494	0.600	1.252	0.586	*Patel et al 2017*	4
137-Cs	3.494	0.800	1.162	0.779	*Guan et al 2015*	4
137-Cs	3.494	9.200	1.770	0.414	*Patel et al 2017*	4
137-Cs	3.494	6.800	1.513	0.759	*Patel et al 2017*	4
137-Cs	3.494	3.900	1.770	0.000	*Patel et al 2017*	4
137-Cs	3.494	2.100	1.431	0.379	*Patel et al 2017*	4
137-Cs	3.494	5.700	1.736	0.000	*Patel et al 2017*	4
137Cs	3.514	2.700	0.735	2.385	*Schuff et al 2002*	4
60-Co	28.467	1.850	1.390	0.852	*Fuhrman Conti et al 1988*	4
137-Cs	1.220	18.600	2.576	7.200	*Guan et al 2015*	5
137-Cs	3.494	18.600	3.394	3.045	*Guan et al 2015*	5
137-Cs	3.494	17.300	2.824	2.055	*Guan et al 2015*	5
137-Cs	3.494	13.200	2.140	0.586	*Patel et al 2017*	5
137-Cs	3.494	19.100	3.026	1.241	*Patel et al 2017*	5
137-Cs	3.494	16.100	2.430	0.621	*Patel et al 2017*	5
137-Cs	3.494	14.800	2.328	0.414	*Patel et al 2017*	5
137-Cs	3.494	17.700	2.733	0.793	*Patel et al 2017*	5
137-Cs	3.494	11.700	1.964	0.345	*Patel et al 2017*	5
137-Cs	3.494	14.800	2.027	1.538	*Guan et al 2015*	5
60-Co	17.100	1.040	1.517	1.749	*Petrovic et al 2006*	Excluded
60-Co	69.500	7.700	1.732	0.230	*Baggio et al 2002*	Excluded
137-Cs	0.051	2.290	0.751	35.000	*Schuff et al 2002*	Excluded
137-Cs	0.051	13.600	0.877	80.000	*Schuff et al 2002*	Excluded

4.2 Cluster Analysis

The lowest AIC-value obtained was at five clusters as seen in Figure 8, and this number of clusters also gave a satisfactory Silhouette measure of 0.46. The BIC value, however, was lowest at three clusters with almost no change between two and four clusters, implying a difference of opinion between the information loss models. The silhouette score did not change much when the number of clusters was altered, however, they

appear to improve slightly for a higher number of clusters. Too many clusters, however, increased the information loss. Because of this, a fixed value of five clusters was chosen for further analysis. This was also the number of clusters chosen by SPSS if a fixed value was not specified. The Silhouette measure of 0.46, is according to SPSS a "fair" clustering, and close to 0.5 being classified as "good".

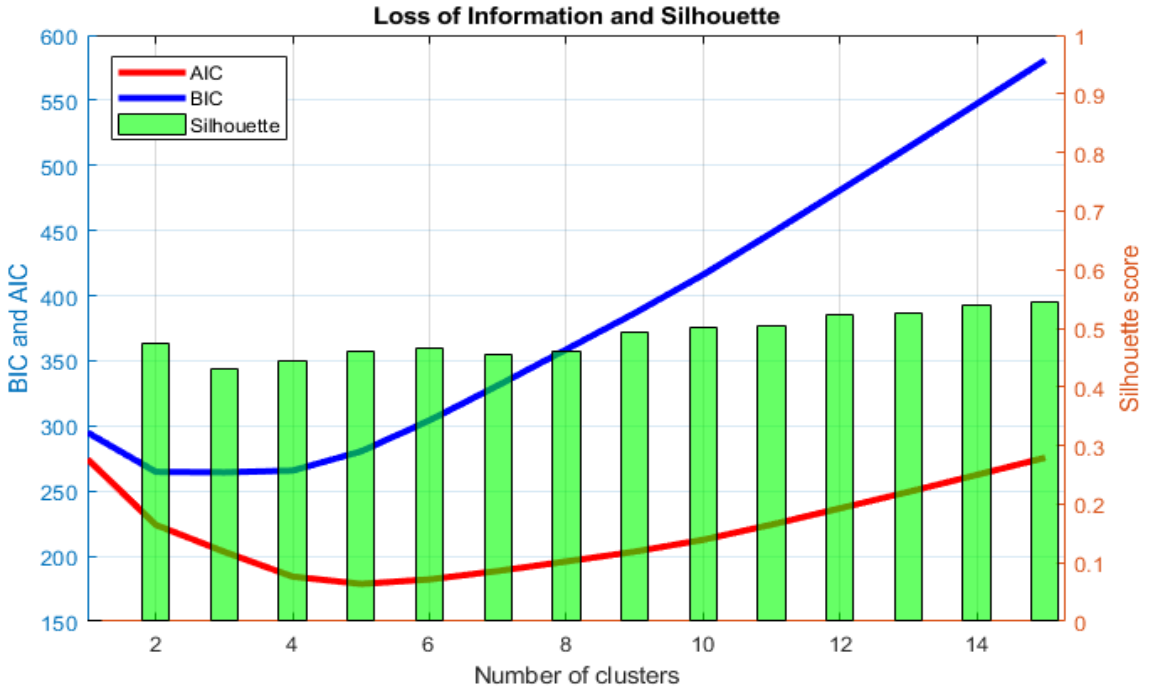

Figure 8: The loss of information as a function of the number of clusters, calculated using Bayesian Information Criterion (Blue) and Akaike Information Criterion (Red) respectively. The Silhouette scores for 2-15 clusters are also depicted as Green bars.

The highest weighted variable when deciding the cluster memberships for each experiment, according to the predictor importance, was the β value closely followed by the D_{min} value at 0.945. Slightly lower weighted was the α value at 0.692. And lowest weighted of the input variables was the LET_d values, weighted at 0.661 relative to β according to equation (21). This implies that the β and D_{min} variables explain the separation of the clusters better than the LET_d and α values can. This can be seen in Figure 9. The cluster memberships of each experiment is listed in Table 2.

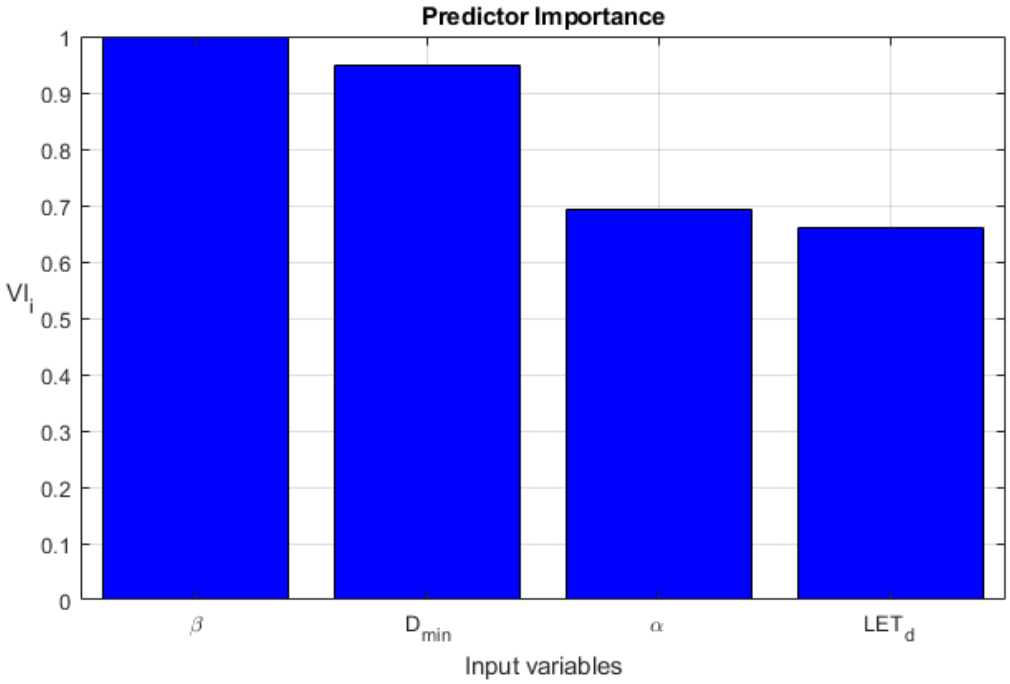

Figure 9: Predictor Importance for each of the input variables, for a 5-cluster model.

The distribution, average values, and standard deviation of each input variable, and for each cluster, are displayed in Figure 10. The same values are displayed for the evaluation fields in Figure 11. The average values and standard deviations are listed in Table 3. The clusters are plotted from bottom to top sorted by increasing average LET_d value within each of the subplots in Figure 10 and Figure 11, meaning that in each subplot the data is sorted from bottom to top in the following order: Cluster 4 (Green), Cluster 1 (Red), Cluster 2 (Blue), Cluster 3 (Purple) and Cluster 5 (Cyan). The subplots themselves are displayed in order of decreasing Predictor Importance from top to bottom.

The number of experiments in each of the five clusters were unevenly distributed into three large clusters (number of datapoints): Clusters 1 (30), 3 (22) and 4 (24) and two smaller clusters: Clusters 2 (8) and 5 (10)).

Table 3: Mean value of each variable considered in the analysis with accompanying standard deviation for each model and for the entire dataset (All clusters), rounded to two decimals.

Mean and Standard deviation						
	Cluster 1	***Cluster 2***	***Cluster 3***	***Cluster 4***	***Cluster 5***	***All Clusters***
$\boldsymbol{\beta}[\boldsymbol{Gy^{-2}}]$	0.05 ± 0.04	0.05 ± 0.04	0.05 ± 0.04	0.11 ± 0.08	0.53 ± 0.22	0.12 ± 0.17
$\boldsymbol{D_{min}}[\boldsymbol{Gy}]$	1.26 ± 0.36	0.28 ± 0.14	0.57 ± 0.26	0.35 ± 0.15	0.33 ± 0.18	0.68 ± 0.48
$\boldsymbol{\alpha}[\boldsymbol{Gy^{-1}}]$	0.30 ± 0.22	1.13 ± 0.45	0.31 ± 0.12	0.14 ± 0.09	0.34 ± 0.25	0.34 ± 0.32
$\boldsymbol{LET_d}\left[\frac{\boldsymbol{keV}}{\boldsymbol{\mu m}}\right]$	5.26 ± 4.58	9.63 ± 6.53	12.89 ± 4.34	2.59 ± 2.20	16.19 ± 2.50	7.90 ± 6.24
$\left(\frac{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\right)_{\boldsymbol{x}}[\boldsymbol{Gy}]$	5.66 ± 4.95	10.33 ± 5.46	6.36 ± 6.07	3.78 ± 5.37	3.27 ± 0.72	5.49 ± 5.38
$\boldsymbol{RBE_{max}}$	1.96 ± 1.25	2.81 ± 1.55	2.27 ± 1.33	0.99 ± 0.74	1.78 ± 2.08	1.84 ± 1.39
$\boldsymbol{RBE_{min}}$	1.16 ± 0.51	0.74 ± 0.52	1.07 ± 0.51	1.19 ± 0.32	2.54 ± 0.46	1.26 ± 0.65

From Figure 10 it can be seen that for α, β and D_{min}, a single cluster contains the distinctly highest values for each of the aforementioned variables, distinctly separated from the other clusters. LET_d however, has a smoother transition from each cluster to the next where the average LET_d increases approximately linearly from one cluster to the next (Figure 10). The standard deviation for the LET_d is also significantly overlapping for the clusters with the closest average LET_d value, however, for the rest of the clusters, the standard deviation is observed overlapping to a decidedly lesser degree.

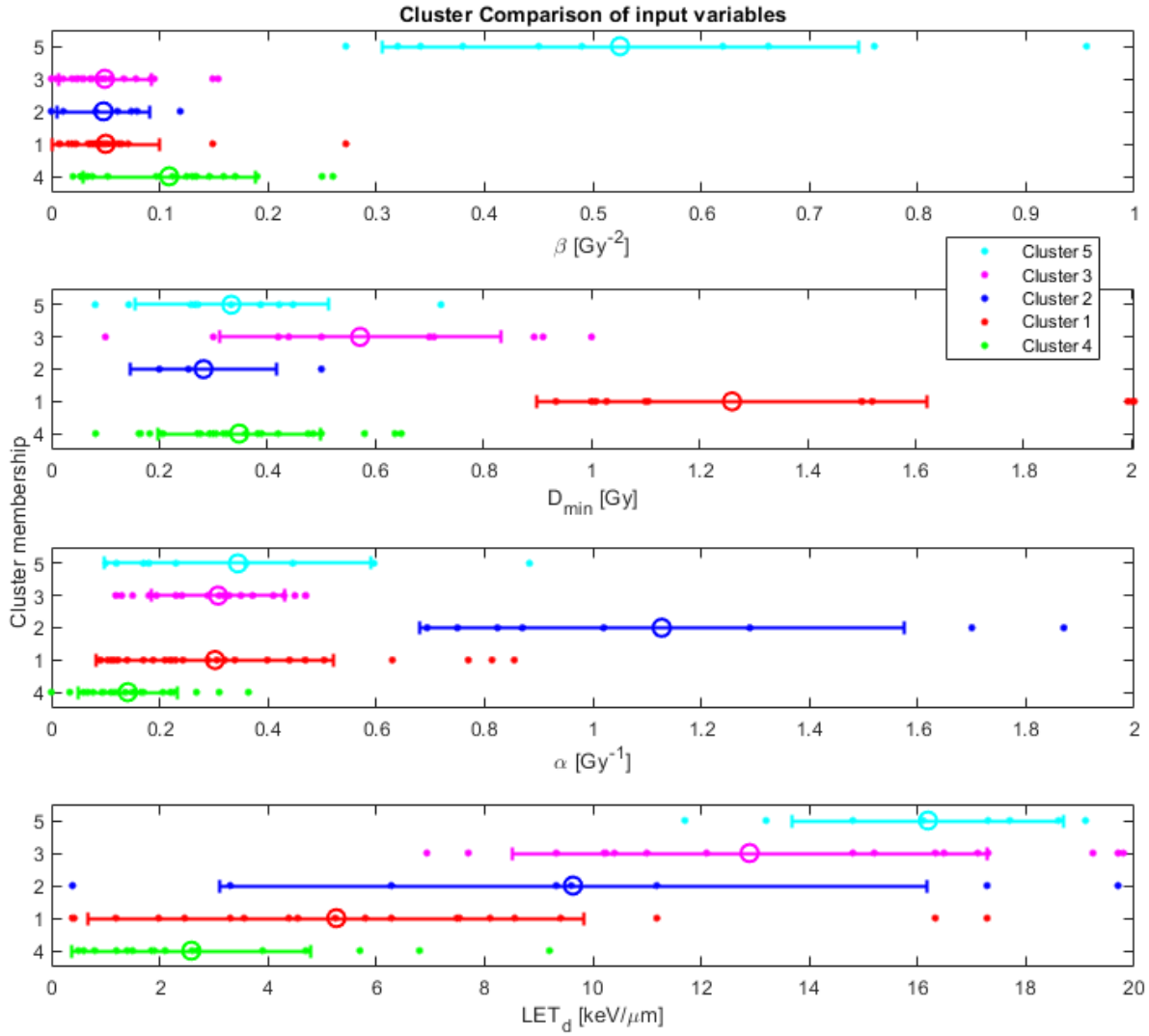

Figure 10: Cluster comparison of input variables. All datapoints in each cluster are shown as dots for each input variable, with mean value marked as a circle with error bars of $\pm$ one standard deviation of the mean. Input variables are listed from top to bottom by decreasing Predictor Importance.

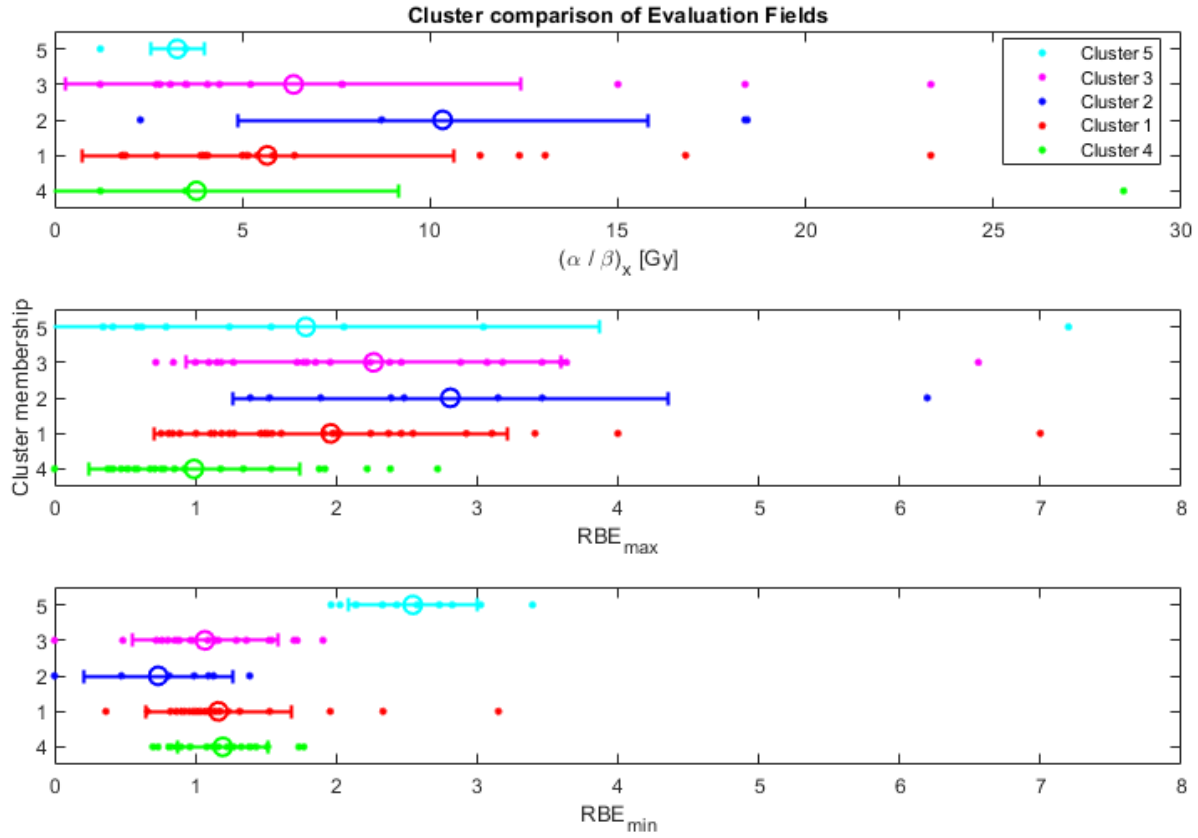

Figure 11: Cluster comparison of evaluation fields. All datapoints in each cluster are shown as dots for each input variable, with mean value marked as a circle with error-bars of ± one standard deviation of the mean.

For the evaluation fields in Figure 11, it can be seen that the only cluster with a standard deviation not overlapping with another cluster is Cluster 5 for the RBE_{min}. This can be explained by the similar form of the β values as seen in Figure 10, and the definition of RBE_{min} as given by equation (10). RBE_{max} has the highest average values in Cluster 2, however, the distribution is not as extreme as the RBE_{min} distribution compared to the other clusters, even though the α distribution for Cluster 2 is similar to the β distribution for Cluster 5.

We can also see that the RBE_{max} and $\left(\frac{\alpha}{\beta}\right)_x$ have a similar shape where they seem to increase for the low LET_d clusters and fall off for the high LET_d clusters. The fact that RBE_{max} and $\left(\frac{\alpha}{\beta}\right)_x$ have the same shape, also indicates that RBE_{max} should be

increasing with increasing $\left(\frac{\alpha}{\beta}\right)_x$. The RBE_{min} seem to have the opposite trend where the average RBE_{min} value decreases for the low LET_d clusters and starts increasing for high LET_d clusters.

Otherwise, overlapping standard deviations in the evaluation fields for all variables are observed to a high degree, where only Cluster 5 has no overlap with the other clusters for RBE_{min}, while for $\left(\frac{\alpha}{\beta}\right)_x$, clusters 5 and 2 avoid overlapping standard deviations. For RBE_{max} every single cluster is observed to have an overlapping standard deviation with the other clusters.

Seven of the experiments with higher RBE_{min}than RBE_{max} are found in Cluster 1, Cluster 2 have none, five are found in Cluster 3, 16 in Cluster 4 and nine in Cluster 5. This is also reflected in Figure 11, where it is clear that Clusters 4 and 5 have a higher average RBE_{min} value than RBE_{max}. This is also supported by the β distribution seen in Figure 10, showing a higher average β value for Cluster 4 and, particularly, Cluster 5, compared to the other clusters, as well as noticeably low α values, especially for Cluster 4. This together with the α and β dependencies of equations (9) and (10) explain the distribution of experiments with higher RBE_{min} than RBE_{max}. The exact opposite is observed for Cluster 2, which has no RBE_{min} values higher than RBE_{max}. The β values for this cluster are among the lowest whereas the α values are clearly the highest.

Results from the cluster analysis, including outliers, can be found in Appendix B. The biggest differences between the two cluster analysis is that for the 98 experiment database, a singleton cluster was created. This cluster included only the *Petrovic et al* experiment, with a D_{min} value of 8.015 Gy. A preferable predictor importance was also observed in this cluster analysis. Excluding the datapoints was still considered to be the correct decision.

4.3 Model Comparison

The model comparisons are ordered according to the percentage of experiments in their database that were also included in the database used in this thesis. The percentage is shown in parenthesis in each title. The following models had five or fewer common datapoints with our database: *Belli et al*, *Jones et al, Carabe et al, Chen and Ahmad et al*, and *Tilly et al*. Of these, only the *Tilly et al* model was analyzed further because of its high percentage of common datapoints. All model equations for the models analyzed further can be found in Appendix C.

4.3.1 McNamara [81] (15%)

The database used as a basis for this model was the database found in *Paganetti et al* [26], which was also one of three major sources for the database used in this thesis. McNamara had the restriction of $LET_d < 20 \frac{keV}{\mu m}$, which is similar to our restriction, and the McNamara model also had a restriction for $\left(\frac{\alpha}{\beta}\right)_x < 30\ Gy$. The main difference in restrictions between the two databases is our restriction of only including monoenergetic experiments. This leads to only 44 shared experiments between the two databases, out of the total 285 experiments in the McNamara database. 17 of the datapoints are assigned to Cluster 1, 8 datapoints in Cluster 2, 17 points in Cluster 3 and 2 points in Cluster 4. Only Cluster 5 is not represented in McNamara's database. The McNamara model predicts an increasing RBE_{max} with increasing LET_d, as well as a steeper slope for low $\left(\frac{\alpha}{\beta}\right)_x$. The model also seems to predict a slight decrease in RBE_{min} with increasing LET_d, whereas $\left(\frac{\alpha}{\beta}\right)_x$ does not appear to have much impact on the slope.

4.3.2 Mairani [82] (42%)

The Mairani database is based on the database from the upcoming Wedenberg model, with an additional seven datapoints. None of these seven additional datapoints were included in our database, meaning that the distribution of data in clusters for the Mairani model is identical to the distribution for the Wedenberg model. This database contains data with $\left(\frac{\alpha}{\beta}\right)_x = \infty$. There was a total of 13 common datapoints, one point from Cluster 1, one point in Cluster 2 and 11 points from Cluster 3. Similar to earlier models, this model predicts an increasing RBE_{max} with increasing LET_d, and a decreasing RBE_{min} with an increasing LET_d. The model considers the biological effect of the primary protons, as well as secondary protons, deuterons, tritons, and helium fragments.

4.3.3 Wilkens and Oelfke [28] (42%)

The Wilkens and Oelfke model consists of 19 datapoints where 7 of them correspond to datapoints in Cluster 1, as well as one datapoint belonging to Cluster 3, adding up to a total of 8 common datapoints. The remaining 11 points were not included in our database, due to our restrictions of only including monoenergetic experiments and LET_d values lower than $20\,\frac{keV}{\mu m}$. In this model it is also suggested that α increases with an increasing LET, however, only α values up to $0.8\ Gy^{-1}$ were included in the model. The model also suggests that β is kept stable around $\beta = 0.0298 Gy^{-2}$, however, the spread within the experimental data increases with increasing LET_d. Furthermore, the model predicts a linearly increasing RBE_{max} with increasing LET_d, up to almost $LET_d = 30\,\frac{keV}{\mu m}$, where after this point the experimental data falls off.

4.3.4 Peeler [83] (50%)

The Peeler model consists of 48 datapoints, 24 from *Guan and Bronk et al* [22] and the remaining 24 from a refit of the same data. The original data from *Guan and Bronk et al* [22] was included in this database and distributed with 4 points in Cluster 3, 16 points in Cluster 4 and the remaining 4 points in Cluster 5. Like most other models this model predicts an increasing RBE_{max} with increasing LET_d, however, this model also predicts an increasing RBE_{min} with increasing LET_d.

4.3.5 Wedenberg [79] (54%)

The Wedenberg model included 24 datapoints, as mentioned in the Mairani comparison in 4.3.2, where five of them came from cells with $\left(\frac{\alpha}{\beta}\right)_x = \infty$. No data from our database shared this quality. From the remaining 19 points, one corresponds to data in Cluster 1, one in Cluster 2 and 11 datapoints can be found in Cluster 5, adding up to a total of 13 common datapoints. The remaining six datapoints were not included in our database. Notably, the lowest LET_d value used in this model was $LET_d = 7{,}7\,\frac{keV}{\mu m}$. This model predicts an increasing RBE_{max} with increasing LET_d, and with a steeper slope for low $\left(\frac{\alpha}{\beta}\right)_x$ values.

4.3.6 Rørvik [25] (71%)

The Rørvik database was one of three used as basis for the database gathered in this thesis, however, with a couple of differences in the restrictions set. Rørvik had restrictions of $LET_d < 40\,\frac{keV}{\mu m}$ and $\left(\frac{\alpha}{\beta}\right)_x < 25\ Gy$, whereas our database restricted LET_d to $20\ keV/\mu m$. Of the 85 datapoints in this database, 60 of them were included in our

database. 11 in Cluster 1, 7 in Cluster 2, 21 in Cluster 3, 17 in Cluster 4 and 4 in Cluster 5.

Rørvik developed two models to fit the data, one weighted and one unweighted linear model. The weighted model depends on a Biological Weighting Function, but for comparison with the linear models created for each cluster, only the unweighted linear model is of interest.

4.3.7 Tilly [67] (81%)

Tilly created two different models where the $\left(\frac{\alpha}{\beta}\right)_x$ was the differentiator. One model consists of 7 datapoints with $\left(\frac{\alpha}{\beta}\right)_x \sim 2\ Gy$. Of these datapoints, one can be found in Cluster 1 and 4 are found in Cluster 3. The remaining 2 datapoints were not included in our database. This model predicts, similarly to many other models, an increasing RBE_{max} with increasing LET_d up to about $LET_d = 30\frac{keV}{\mu m}$ with an experimental decrease of RBE_{max} after this point. The other model also includes 4 datapoints with $\left(\frac{\alpha}{\beta}\right)_x \sim 10\ Gy$. All the aforementioned datapoints can be found in Cluster 3. This model also predicts a rise in RBE_{max} with increasing LET_d, however, with a significantly gentler slope than the model with lower $\left(\frac{\alpha}{\beta}\right)_x$.

4.4 RBE modelling

The curve fit tool from MATLAB gave us fitting variables for the RBE_{min} equations in the range of $-0.031\ \left(\frac{keV}{\mu m}\right)^{-1}$ to $0.097\ \left(\frac{keV}{\mu m}\right)^{-1}$. For the RBE_{max} curve fit we got fitting variables in the range $-0.006\ Gy\left(\frac{keV}{\mu m}\right)^{-1}$ to $1.271\ Gy\left(\frac{keV}{\mu m}\right)^{-1}$, showing a significantly greater spread.

The fitting variable, the R-squared value, and the 95% confidence interval for each model are listed for the RBE_{min} curve fit in Table 4, and the same are listed for the RBE_{max} curve fit in Table 5.

A positive fitting variable predicts an increasing RBE with increasing LET_d for both RBE_{min} and RBE_{max}, given a set $\left(\frac{\alpha}{\beta}\right)_x$ value for the RBE_{max}. This can be seen in Figure 12 and Figure 13.

Table 4: Fitting variable for RBE_{min}, with accompanying goodness of fit statistic R-squared, and 95% confidence interval for each of the models created with the curve fit tool in MATLAB.

RBE_{min} curve fit variable			
Model	$\boldsymbol{c_{min}}\left[\left(\frac{\boldsymbol{keV}}{\boldsymbol{\mu m}}\right)^{-1}\right]$	***R-squared***	***95% confidence interval***
Cluster 1	0.027	0.030	$[-0.001, 0.054]$
Cluster 2	-0.031	0.223	$[-0.065, 0.003]$
Cluster 3	0.008	0.034	$[-0.008, 0.025]$
Cluster 4	0.082	0.405	$[0.052, 0.113]$
Cluster 5	0.097	0.633	$[0.085, 0.109]$
All clusters	0.034	0.118	$[0.021, 0.046]$

Table 5: Fitting variable for RBE_{max}, with accompanying goodness of fit statistic R-squared, and 95% confidence interval for each of the models created with the curve fit tool in MATLAB.

RBE_{max} *curve fit variable*			
Model	$c_{max}\left[Gy\left(\frac{keV}{\mu m}\right)^{-1}\right]$	***R-squared***	***95% confidence interval***
Cluster 1	0.583	0.710	[0.479, 0.687]
Cluster 2	1.271	0.870	[1.008, 1.534]
Cluster 3	0.303	0.615	[0.235, 0.371]
Cluster 4	−0.006	$-5.488 * 10^{-5}$	[−0.237, 0.225]
Cluster 5	0.248	0.512	[0.088, 0.408]
All clusters	0.331	0.426	[0.273, 0.389]

The fitted curves for RBE_{min} are shown in Figure 12, and the fitted planes for RBE_{max} are shown in Figure 13.

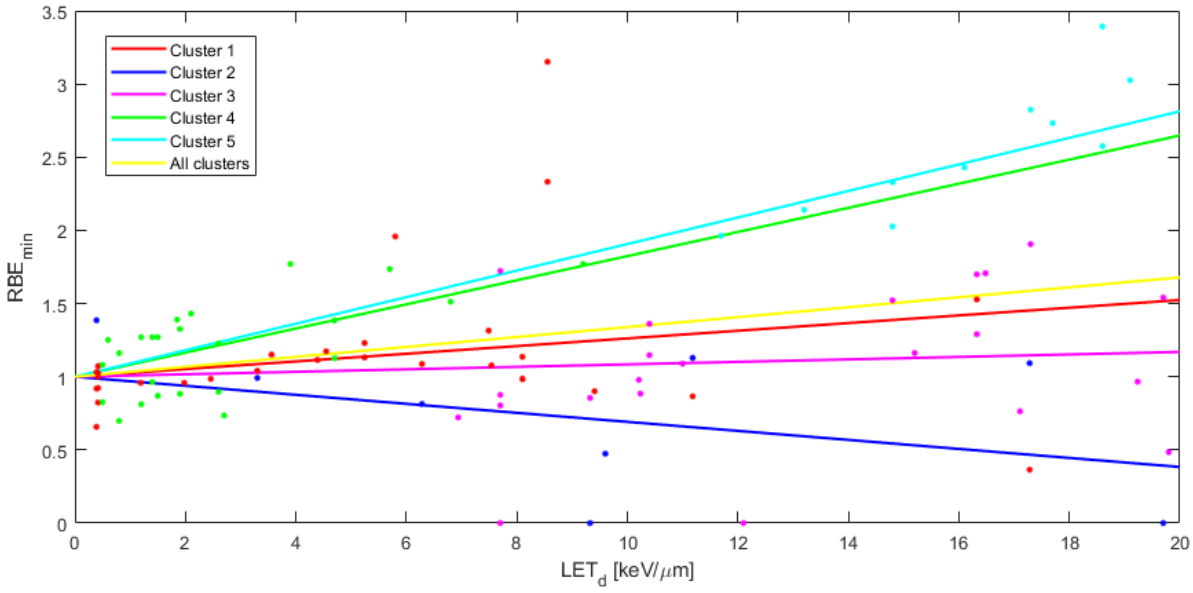

Figure 12: Each line depicts the RBE_{min} as a function LET_d by fitting the experimental data in each cluster to equation (22). The yellow line represents the model where all experimental data in all clusters was fitted to the same equation.

Clusters 4 and 5 predict a slope for the RBE_{min} of about 0.1 with increasing LET_d, as seen by their fitting variables, whereas Clusters 1 and 3 predict a significantly shallower slope. Finally, Cluster 2 is the only Cluster to predict a decreasing RBE_{min} to LET_d, with a slope of -0.03. As expected, the model considering all the data is located close to the centre of the other models.

For the linear RBE_{min} fits, we observe a wide spread in the experimental data represented in the low R-squared values in Table 4. The Cluster 1 and Cluster 3 models have the lowest R-squared values of 0.034 or lower, whereas the Cluster 4 and Cluster 5 models present fewer variations with higher R-squared values of 0.40 and 0.63, respectively. The Cluster 2 model, and the All clusters model present R-squared values of 0.22 and 0.12, respectively, representing a bad fit. They are, however, not as bad as the Cluster 1 and Cluster 3 models.

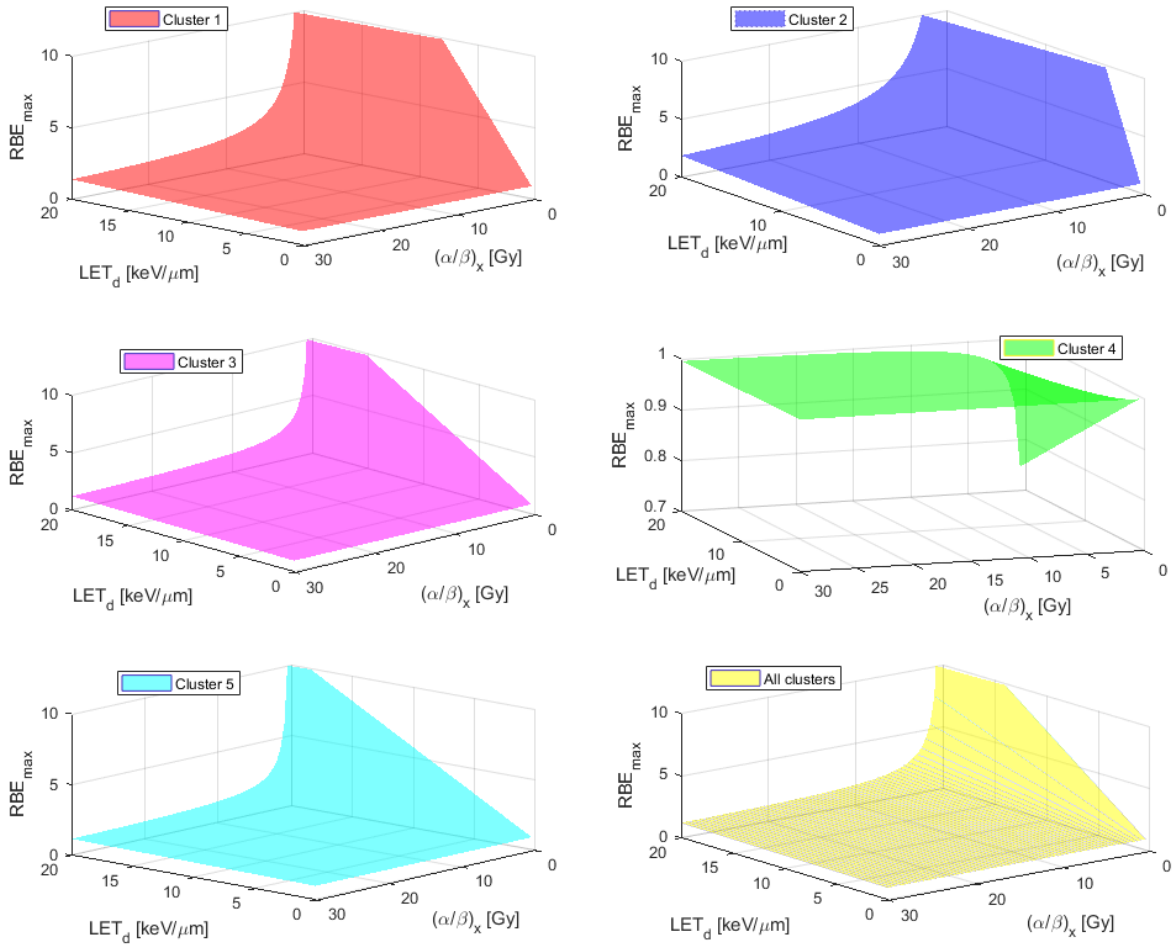

Figure 13: Each subplot depicts the RBE_{max} as a function of LET_d and $\left(\frac{\alpha}{\beta}\right)_x$ by fitting the experimental data from each cluster to equation (23). The yellow plane represents the model where all experimental data in all clusters was fitted to the same equation.

From Figure 13 it is clear that most models predict an increasing RBE_{max} with increasing LET_d, where the slope gets steeper for lower $\left(\frac{\alpha}{\beta}\right)_x$ values. Only the model from Cluster 4 predicts a decreasing RBE_{max} with increasing LET_d, and the slope gets steeper for lower $\left(\frac{\alpha}{\beta}\right)_x$ values in this model as well, albeit with a negative slope. Clusters 1 and 2 predict the steepest slope with regards to LET_d.

The overall "goodness of fit" within the RBE_{max} fit is significantly better than within the RBE_{min} fit. This is seen from the high R-squared values as presented in Table 5. The obvious exception is the Cluster 4 model with a negative R-squared value. This is,

according to MATLAB only possible "if the model does not contain a constant term and the fit is worse than just fitting the mean". The rest of the models have R-squared values between 0.426 and 0.870, where the Cluster 2 model represents the best fit to its experimental data, and the All clusters model represents the worst fit, if the Cluster 4 model is disregarded. The fit of the best and worst fitted planes to its models experimental data (Cluster 2 and Cluster 4, respectively) are visualized in Figure 14.

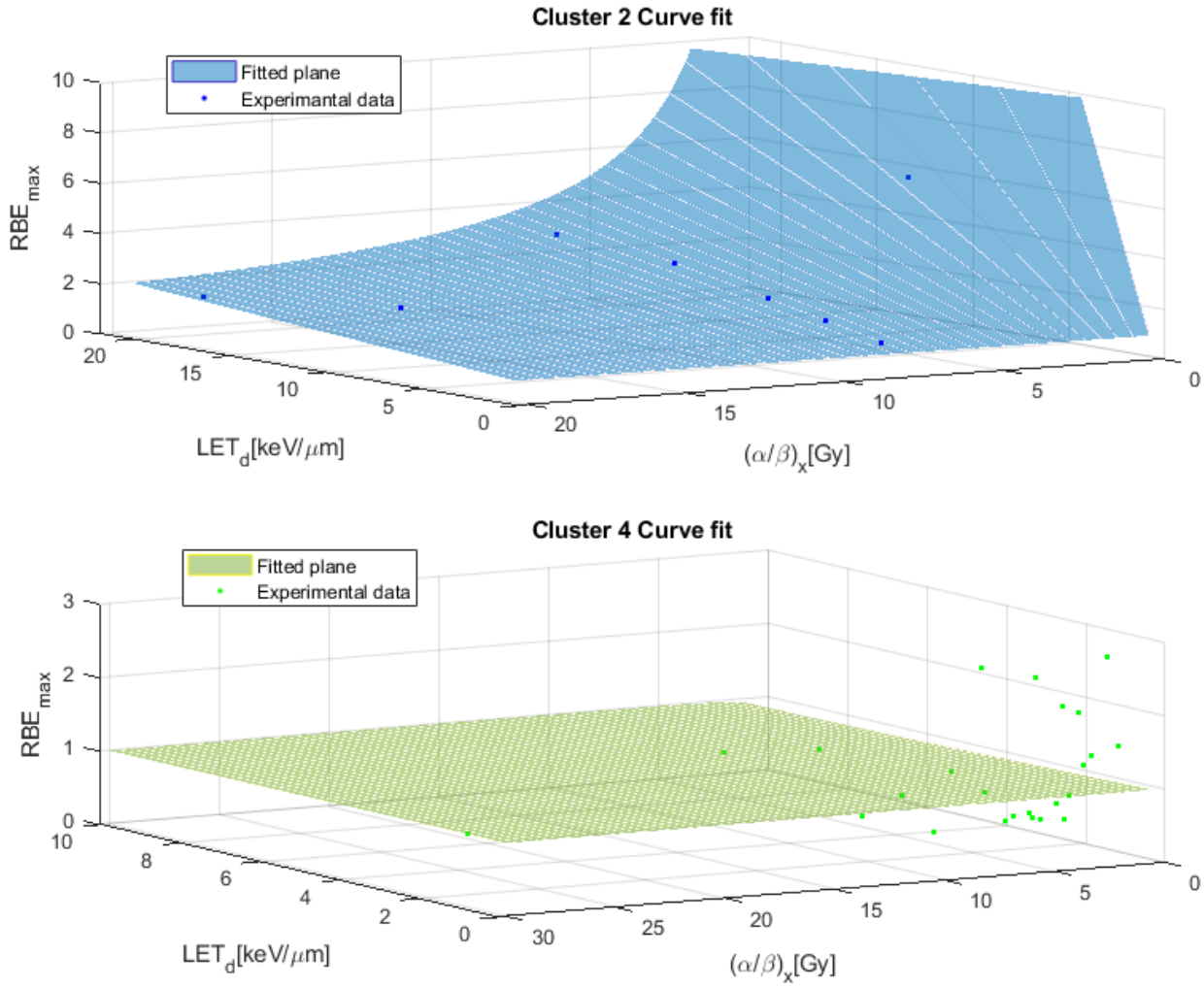

Figure 14: Fitted plane representing the RBE_{max} for Cluster 2 and Cluster 4 plotted together with the experimental data from the same clusters to illustrate the "goodness of fit". A slightly different shade, of blue and green respectively, was utilized in this plot to make the experimental datapoints easier to see. The z-axis for Cluster 4 in the plot is larger than the one in Figure 13 to make room for all the experimental data. This also makes the slope when $\left(\frac{\alpha}{\beta}\right)_x \rightarrow 0$ less clear.

The model equations for each of the models are written explicitly as follows:

Cluster 1

$$RBE_{min}(LET_d) = 1 + 0.027\left(\frac{keV}{\mu m}\right)^{-1} LET_d \tag{24}$$

$$RBE_{max}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 1 + \frac{0.583Gy\left(\frac{keV}{\mu m}\right)^{-1}}{\left(\frac{\alpha}{\beta}\right)_x} LET_d \tag{25}$$

Cluster 2

$$RBE_{min}(LET_d) = 1 - 0.031\left(\frac{keV}{\mu m}\right)^{-1} LET_d \tag{26}$$

$$RBE_{max}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 1 + \frac{1.271Gy\left(\frac{keV}{\mu m}\right)^{-1}}{\left(\frac{\alpha}{\beta}\right)_x} LET_d \tag{27}$$

Cluster 3

$$RBE_{min}(LET_d) = 1 + 0.008\left(\frac{keV}{\mu m}\right)^{-1} LET_d \tag{28}$$

$$RBE_{max}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 1 + \frac{0.302Gy\left(\frac{keV}{\mu m}\right)^{-1}}{\left(\frac{\alpha}{\beta}\right)_x} LET_d \tag{29}$$

Cluster 4

$$RBE_{min}(LET_d) = 1 + 0.082\left(\frac{keV}{\mu m}\right)^{-1} LET_d \quad (30)$$

$$RBE_{max}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 1 - \frac{0.006Gy\left(\frac{keV}{\mu m}\right)^{-1}}{\left(\frac{\alpha}{\beta}\right)_x} LET_d \quad (31)$$

Cluster 5

$$RBE_{min}(LET_d) = 1 + 0.097\left(\frac{keV}{\mu m}\right)^{-1} LET_d \quad (32)$$

$$RBE_{max}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 1 + \frac{0.248Gy\left(\frac{keV}{\mu m}\right)^{-1}}{\left(\frac{\alpha}{\beta}\right)_x} LET_d \quad (33)$$

All clusters

$$RBE_{min}(LET_d) = 1 + 0.034\left(\frac{keV}{\mu m}\right)^{-1} LET_d \quad (34)$$

$$RBE_{max}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 1 + \frac{0.331Gy\left(\frac{keV}{\mu m}\right)^{-1}}{\left(\frac{\alpha}{\beta}\right)_x} LET_d \quad (35)$$

4.5 FLUKA simulation and dose calculatons

The two-dimensional dose distribution, calculated based on the FLUKA MC simulations in Chapter 3.5, is shown in Figure 16. It depicts the distribution for all the relevant models in a slice 1 mm from the center of the phantom, as the voxel file definitions made this the closest slice to the center of the PTV. A one-dimensional depth dose distribution for the same models is also shown in Figure 15. These figures show that only the $RBE = 1.1$ model predicts a reasonably homogeneous dose distribution, which is explained by the model not taking the LET_d into account. The rest of the models predict a higher dose in the distal edge of the SOBP, where the models for Cluster 1 and Cluster 2 clearly predict the highest doses. This is explained by the high fitting variables for the RBE_{max} model equations (25) and (27), for these models in particular. The LET_d distribution, obtained from the FLUKA simulation shows an increasing LET_d in the distal edge of the SOBP and is given in Appendix E.

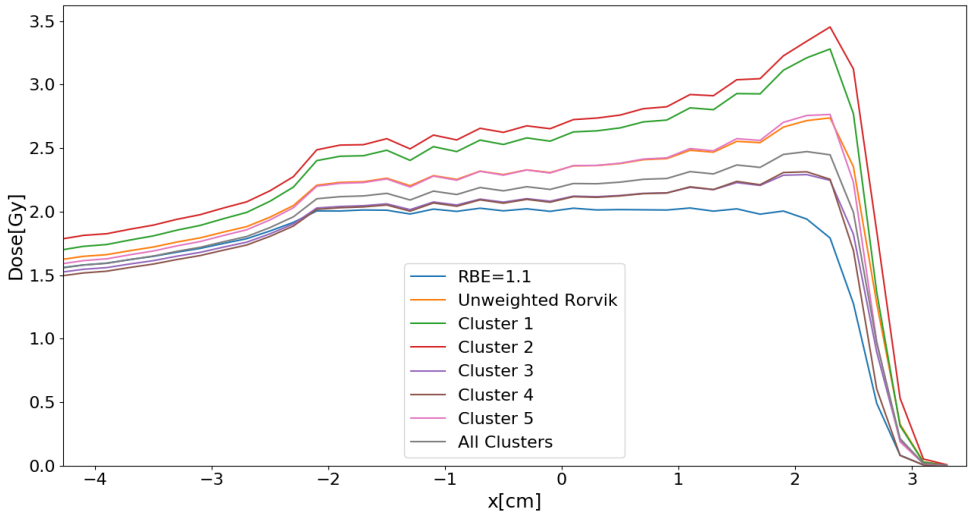

Figure 15: Depth dose distribution of the simulated proton beam calculated for RBE of 1.1, Unweighted Rørvik model (Chapter 4.3.6)), models for each clusters and model based on all clusters (Chapter 4.4).

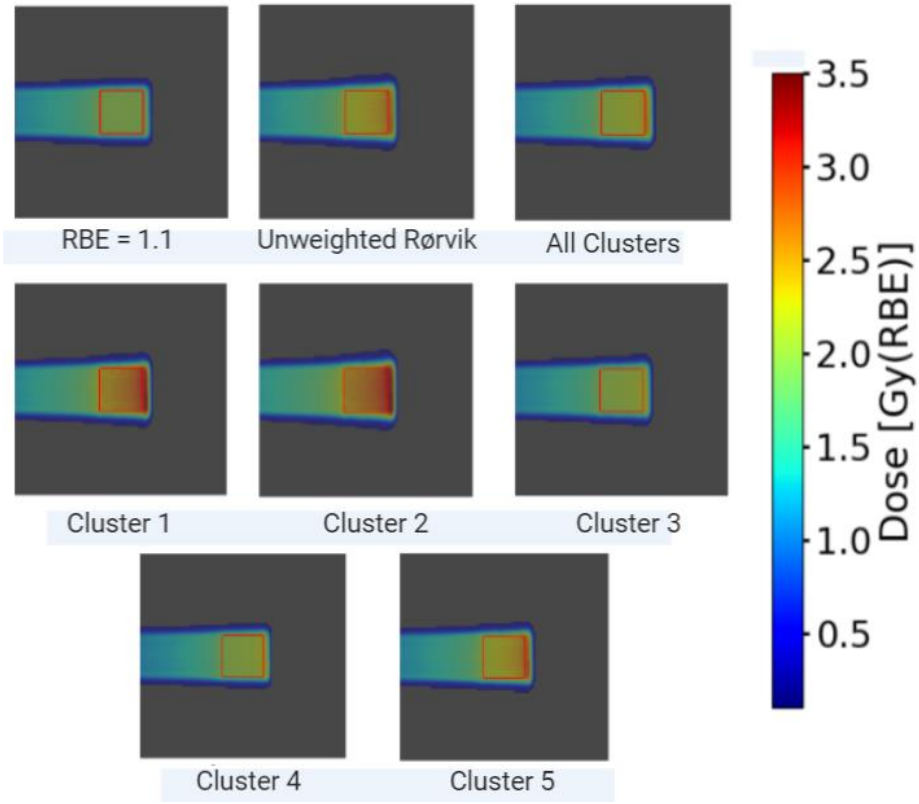

Figure 16: Dose distribution based on the FLUKA Monte Carlo simulation calculated with RBE of 1.1, Unweighted Rørvik model [25] equations (Chapter 4.3.6), model equations for all clusters and model equations for each separate cluster (Chapter 4.4). The PTV is illustrated as a red square.

The DVHs from the MC simulation are shown in Figure 17. It is again clear that the models for Cluster 1 and Cluster 2 predict a distinctly higher dose to the PTV, whereas the $RBE = 1.1$ model gives the most homogenous and lowest dose. The maximum and minimum doses, as well as the dose absorbed by at least 50% of the PTV is listed for each model in Table 6. It is shown here that voxels within the PTV receive absorbed doses in the range 1.6 $Gy(RBE)$ to 3.5 $Gy(RBE)$. From Figure 15, Figure 16, Figure 17, and Table 6 we observe a striking resemblance between the Cluster 5 model and the unweighted Rørvik model, with an almost identical dose distribution and DVH values.

Table 6: DVH values of $D(0\%)$ or max dose, $D(50\%)$ measuring the those absorbed by 50% of the PTV, and $D(100\%)$ or minimum dose for the PTV.

Model	$D_{max}[Gy(RBE)]$	$D_{50\%}$ $[Gy(RBE)]$	D_{min} $[Gy(RBE)]$
$RBE_{1.1}$	2.10	1.97	1.60
Unweighted Rørvik	2.80	2.31	1.70
All Clusters	2.60	2.18	1.70
Cluster 1	3.30	2.56	1.90
Cluster 2	3.50	2.67	2.00
Cluster 3	2.40	2.08	1.60
Cluster 4	2.40	2.07	1.60
Cluster 5	2.90	2.31	1.70

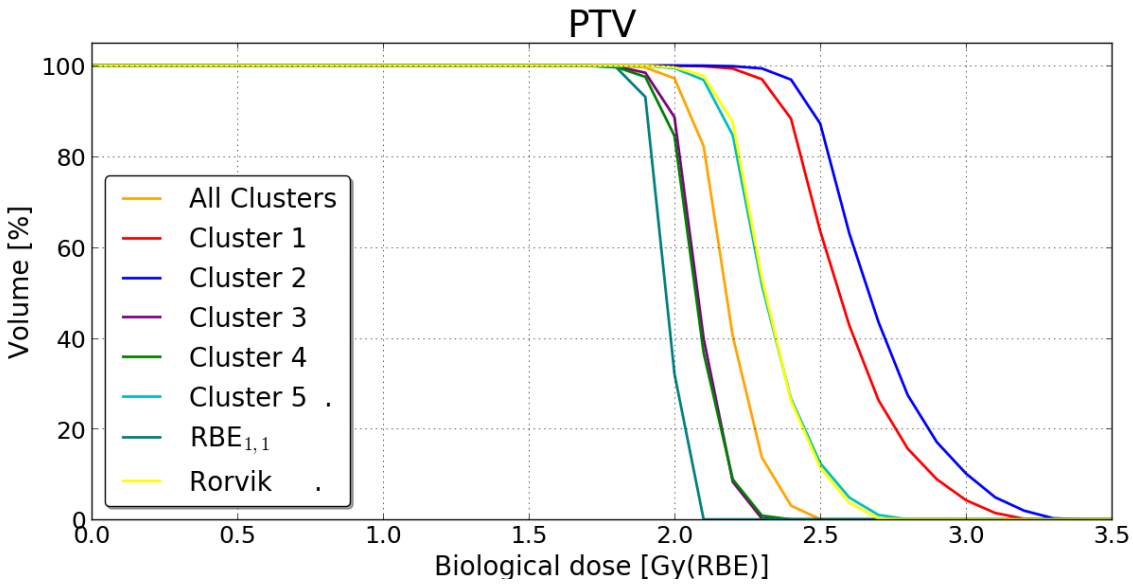

Figure 17: The DVH of biological dose absorbed by the PTV for all models.

The RBE distributions are shown in Figure 18Figure 18. In this plot, voxels with a dose value lower than 0.1 $Gy(RBE)$ are set transparent. The RBE distribution, similarly to the dose distribution, seem to be highest in the distal edge of the field, where again the Cluster 1 and Cluster 2 models predict the highest RBE of the cluster models. The Cluster 1 model predicts a very similar RBE distribution to the unweighted Rørvik model.

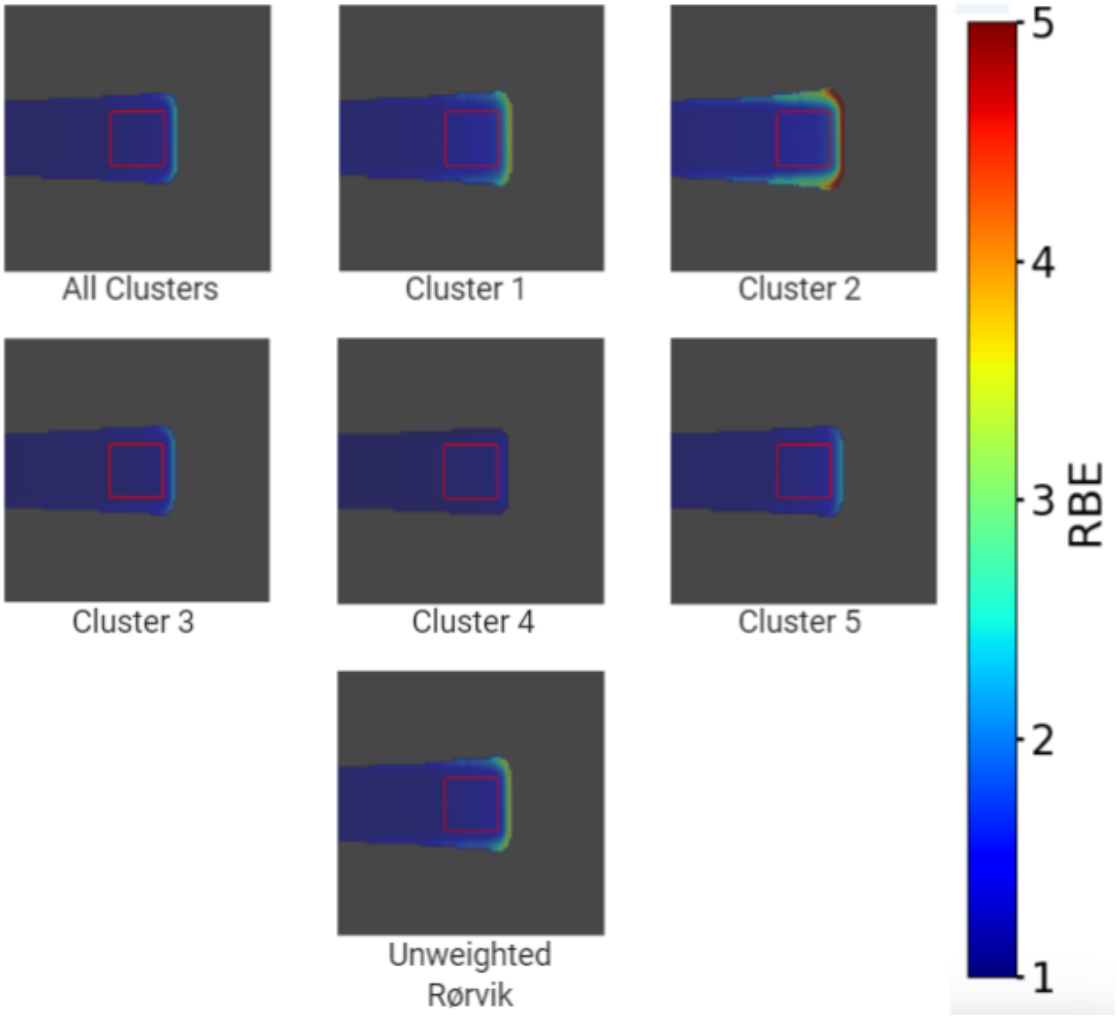

Figure 18: Two-dimensional RBE distribution from the FLUKA simulation recalculated from the model equations created in this thesis with the unweighted Rørvik model for comparison. Voxels that absorbed a dose lower than $0.1\ Gy$ were set transparent.

5. Discussion

From our cluster analysis, five clusters (sub-databases) were created. For three of the input variables used in the analysis, α, β, and D_{min}, one cluster distinguished itself from the others within each of the three variables, showing a clearly higher average value from the other clusters. These clusters were Cluster 2 for the α values, Cluster 5 for the β values, and Cluster 1 for the D_{min} values, respectively, as seen in Figure 10. Common RBE-modelling relationships for RBE_{min} and RBE_{max} were investigated for each cluster by fitting them to equations (22) and (23). The fitting variables for the RBE_{min}, representing the slope as a function of LET_d, had only limited variation ranging from -0.031 $\left(\frac{keV}{\mu m}\right)^{-1}$ to 0.097 $\left(\frac{keV}{\mu m}\right)^{-1}$, where only Cluster 2 predicts a negative slope. The fitting variables for the RBE_{max}, representing the slope as a function of $\frac{LET_d}{\left(\frac{\alpha}{\beta}\right)_x}$, was found in the range of -0.006 $Gy\left(\frac{kev}{\mu m}\right)^{-1}$ to 1.271 $Gy\left(\frac{keV}{\mu m}\right)^{-1}$. Most models reviewed in this thesis assumed a $RBE_{min} = 1$. For RBE_{max}, however, we can see that most models found in Appendix C have a fitting variable closest to the "All clusters" model emulated in this thesis. The McNamara, Mairani, and Wedenberg models all have a RBE_{max} fitting variable close to the one found for the All clusters model. The unweighted Røvik model, and the Tilly model for $\left(\frac{\alpha}{\beta}\right)_x \sim 10Gy$, also depict a fitting variable close to the Clulster 1 fitting variable. This is found by comparing the values in Table 5 with the equations found in Appendix C. Only Cluster 4 predicts a negative slope for the RBE_{max}. RBE_{min}, generally, displays a greater variation in the relevant experimental data as reflected by the R-squared values in Table 4.

The clusters, and their model relationship, were also explored for recalculation of the biological dose and RBE distributions from a simulated proton irradiation. The high α, and high D_{min} databases of the Cluster 1 and Cluster 2 models resulted in the distinctly highest doses to the distal edge of the simulated SOBP, and the clearly highest DVH values as illustrated in Figure 15, Figure 16, and Figure 17.

5.1 Database

The fact that nearly 38% of the experiments in the final database resulted in a higher RBE_{min} value than RBE_{max} value is very surprising and contradicts the outcome expected from the theory. This can possibly be explained by the IndRep and LQC models in the low and high dose ranges, respectively, as these models propose a correction of the inaccuracies of the LQ model in these dose ranges.

Many of the experiments included in our database are quite recent, such as the ones from *Howard et al* and *Patel et al*, which were both published in 2017. Comparing our database to the databases used in earlier models, such as *Wilkens and Oelfke et al*, and *Tilly et al*, from 2004 and 2005, respectively, is not completely fair as these models would not have access to the newer data. However, there was still large amounts of experimental data available when these models were published.

The restriction of $LET_d < 20 \frac{keV}{\mu m}$ is reasonable as this is the range used clinically. The inclusion of the lowest dose value from each experiment was thought to reflect the preciseness of the RBE_{min}, as this is defined in the low dose region. Earlier studies have shown that D_{min} affect the fit of RBE_{min} [8]. However, by the same logic, the maximum dose value from each experiment could, therefore, also be included, as this is the region where the RBE_{max} is defined. However, as the max dose in most experiments was well above the typical fraction dose of 2 Gy, the impact of the maximum dose may be less than that of D_{min}.

5.1.1 Quality of the cluster analysis

The Silhouette score did not vary much, and the AIC and BIC models did not agree on the optimal number of clusters. This indicates that the clustering might not be optimal, and a different number of clusters could be argumented for. This is also reflected in the sub-optimal Predictor Importance. A model with more even predictor importance would be preferable, such as the one presented in Appendix B. The drop in Predictor

Importance, especially in the variables α and LET_d, means that the clusters can be misrepresented, especially regarding RBE_{max}, as these variables are most commonly used when deriving RBE_{max}, whereas β and D_{min} is commonly not considered as RBE parameters. These quality indicators were considered decent and the cluster analysis was kept, however, a model with a different number of clusters or even another clustering algorithm could possibly present a more representative result.

Including the $\left(\frac{\alpha}{\beta}\right)_x$ as an input variable in the cluster analysis could also be questioned. Even if it could be considered a "swamping variable", meaning a variable that hijacks the clustering, it does present important information on modelling with regards to the tissue, and how it might react to radiation. As mentioned earlier, the maximum dose included in each experiment could also be included in the database, and also as an input variable in the cluster analysis, by the same logic as for the minimum dose. The number of dose – survival fraction datapoints in each experiment may also affect the LQ fit and thereby, the α and β parameters. This could also be introduced as a variable included in the cluster analysis.

5.1.2 Clustering results

The results from the clustering show a similar trend for RBE_{max} and $\left(\frac{\alpha}{\beta}\right)_x$, with the average RBE_{max} and $\left(\frac{\alpha}{\beta}\right)_x$ values increasing from cluster to cluster with increasing average LET_d values, for the lowest LET_d clusters (Clusters 4, 1, and 2, ordered by increasing LET_d). For the high LET_d clusters (Clusters 2, 3, and 5, ordered by increasing LET_d), we observed a decrease in RBE_{max} and $\left(\frac{\alpha}{\beta}\right)_x$ with increasing LET_d. This implies an increasing RBE_{max} and $\left(\frac{\alpha}{\beta}\right)_x$, up to an LET_d value of about 10 $\frac{keV}{\mu m}$, with a following drop for both variables. This can be seen in Figure 10 and Figure 11. A similar trend is reported in several of the models investigated in chapter 3.3, such as *Tilly et al* [67] and *Wilkens and Oelfke* [28], however, in these models a decrease in

RBE_{max} does not occur until LET_d values of about 30 $keV/\mu m$, whereas our database had a maximum LET_d value of 20 $keV/\mu m$. The high standard deviations in both LET_d and RBE_{max} in our study do, however, leave this dependency weak.

For RBE_{min} we found the complete opposite result, compared to that of RBE_{max} and $\left(\frac{\alpha}{\beta}\right)_x$. The average RBE_{min} value decreases with increasing LET_d for the low LET_d clusters. After this, an increase in RBE_{min} is seen from Cluster 2 to Cluster 3 and on to Cluster 5. A decrease in RBE_{min} with increasing LET_d is also reported in the *Mairani et al* [82] and *McNamara et al* [81] models among others, however a following increasing RBE_{min} for higher LET_d values have not been found in any literature studied for this thesis. In this aspect, most RBE models assume a constant RBE_{min} equal to one [31].

The distinct separation in RBE_{min} values for Cluster 5, compared to the other clusters, is clearly explained by the similar separation in β, as this is used to derive the RBE_{min} value mathematically from Equation (10). A similar separation could then be expected to be found for the RBE_{max} value for Cluster 2. Cluster 2 has the highest average RBE_{max}, but there is a considerably smaller degree of separation from the other clusters, when compared with the equivalent separation for the α values. This would probably be a result of similarly high α values in the reference radiation, which could also be implied by Cluster 2 having the highest average $\left(\frac{\alpha}{\beta}\right)_x$ values. However, this requires further investigation to be decisively concluded.

When comparing the final cluster analysis with the previous one, found in Appendix B, the first observation is a slightly better silhouette score and predictor importance, with almost no difference in AIC. We also observe a singleton cluster containing the excluded experiment from *Petrovic et al* [58]. However, if we consider that a singleton cluster is not analytically valuable, and we only consider the other clusters, we observe a clear similarity to the final cluster analysis. For the α and β distribution, we can see that one cluster is separate from the others, implying that Cluster 5 in the final analysis

is similar to Cluster 4 from the old, and the Cluster 2 in the new analysis is similar to Cluster 5 from the old one. A similar comparison of Cluster 1 and Cluster 3 is not observed, however, based on the D_{min} and LET_d distributions, it appears that Cluster 1 and Cluster 3 in the final analysis is a differently divided distribution of the datapoints also found in Cluster 1 and Cluster 3 in the old analysis. We also observe similar trends as reported earlier in the $\left(\frac{\alpha}{\beta}\right)_x$, RBE_{max}, and RBE_{min} distributions.

5.2 RBE modeling

From Figure 11 it can be seen that the RBE_{min} values of all clusters, except for Cluster 5, revolve around the value of 1. Cluster 5, being the cluster separated furthest from the others, due to its high β values, suggests that models with mostly high β values in their database, and hence also tissue with low $\left(\frac{\alpha}{\beta}\right)_x$, also a quality of Cluster 5, should not apply the assumption of RBE_{min} = 1. For models with databases containing mostly lower β values, however, this assumption seems to be valid. This is also reflected with Cluster 5 having the highest fitting variable for the RBE_{min} refit. Cluster 4 only has a slightly lower fitting variable, however, this cluster contained mostly low experimental LET_d values not allowing the model to express the high RBE_{min} values it predicts in the high LET_d range. However, the relatively high variation in this data, as given by the R-squared value, is an argument against $RBE_{min} = 1$. Actually, the clusters augmenting most against $RBE_{min} = 1$, Cluster 4, and Cluster 5, both have the best R-squared values as seen in Table 4. The linear fit to the data was generally poorer for the RBE_{min} fit, than the variation for the RBE_{max} fit, as represented by the lower R-squared values found in Table 4 and Table 5.

The fitting variables for RBE_{max}, however, presents a larger range than the fitting variables for RBE_{min}. The Cluster 2 model, containing the highest fitting variable, presents the steepest RBE_{max} as $\left(\frac{\alpha}{\beta}\right)_x$ decreases despite the model being represented

by the highest average $\left(\frac{\alpha}{\beta}\right)_x$, thus possibly not representing the low $\left(\frac{\alpha}{\beta}\right)_x$ range precisely.

The poor fit to the experimental data of Cluster 4, as seen in Figure 14, seems to be a result of an overload of experimental datapoints with RBE_{max} values lower than one, before increasing with decreasing $\left(\frac{\alpha}{\beta}\right)_x$. Allowing MATLAB to determine the constant term based on the data alone could probably fix this problem.

The size of each cluster is also relevant, as regression analysis often demands a normal distribution in the database to which the regression is applied. Small databases, such as Cluster 2 and Cluster 5, with 8 and 10 datapoints, respectively, can be hard to argument a normal distribution for. With this in mind, RBE models have been created using smaller databases, such as the two Tilly models with four and seven datapoints.

5.3 FLUKA simulation and dose calculations

Common for all dose distributions, recalculated with an LET_d-based RBE model, is an increasing dose to the distal edge of the SOBP, as a result of a distal increase in LET and RBE in the same region. The highest doses calculated with the Cluster 1 and Cluster 2 models separated themselves from the other models with their distinctly high D_{min} and α values, respectively.

The fact that Cluster 5 displays a similar dose distribution to the unweighted Rørvik model is surprising, as the model equations for these models are not very similar. Based on the model equations, it would be expected that the Cluster 1 model would present a dose distribution most similar to the Rørvik model, such as the RBE distribution shows. This may be explained by the higher RBE_{min} fitting variable for Cluster 5.

Based on the dose distributions of the different models, it appears that models based on databases with high α values, and hence high $\left(\frac{\alpha}{\beta}\right)_x$ values, as seen in Figure 10 and

Figure 11, or high D_{min} values, predict a disproportionally high dose to the distal edge of the SOBP. The lowest doses, however, seem to be predicted by models derived from low α, and low LET_d, databases.

5.4 Further work

Given more time, a similar cluster analysis to the one performed on our database would be performed on the original databases for the models included in Chapter 4.3, to investigate if the fitting variables found in these models would differ notably for their clusters. Statistically, this would be more sound, due to a greater number of experiments in each database. It is still unclear which variables should ideally be included in such an analysis. This analysis was not considered in this thesis, since our purpose was to explore potential patterns in a given database for future construction of new RBE-models.

Another option to further investigate the impact of different experimental dose ranges could be to refit the cell survival curves from each experiment in our database, by excluding datapoints outside certain dose ranges. The purpose of this would be to further calculate new α and β values, which would further be implemented in the upper and lower RBE limits.

6. Conclusion

This project applied cluster analysis to investigate relationships between *in vitro* based RBE databases, and the dependencies of the RBE models derived from these data. The results presented in this thesis show that the *in vitro* data can be organized in clusters with different properties, leading to different RBE models and estimates. However, the clusters had some overlap and were distributed relatively evenly over the range of the input variables present in the database. Therefore, the cluster analysis indicates that the database used in this work is relatively well balanced and homogeneous. Nevertheless, it seems like databases overrepresented by high α values, and thus $\left(\frac{\alpha}{\beta}\right)_x$ ratios, predict the highest doses, followed by databases of high D_{min} values. Databases representing low LET_d and α predict the lowest doses.

Bibliography

1. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer J Clin, 2018. **68**(6): p. 394-424.
2. American Cancer Society. *Lifetime Risk of Developing or Dying From Cancer*. 04. January 2018 04. December 2019]; Available from: https://www.cancer.org/cancer/cancer-basics/lifetime-probability-of-developing-or-dying-from-cancer.html.
3. HF, H.B. *Innføring av protonbehandling til Norge*. 2020 [cited 2020 February 21.]; Available from: https://helse-bergen.no/om-oss/protonsenter-i-bergen/innforing-av-protonbehandling-til-norge.
4. Gianfaldoni, S., et al., *An Overview on Radiotherapy: From Its History to Its Current Applications in Dermatology.* Open Access Maced J Med Sci, 2017. **5**(4): p. 521-525.
5. American Society of Clinical Oncology. *Side Effects of Radiation Therapy*. 2018 [cited 2018 December]; Available from: https://www.cancer.net/navigating-cancer-care/how-cancer-treated/radiation-therapy/side-effects-radiation-therapy.
6. National Cancer Institute. *Radiation Therapy Side Effects*. 2018 [cited 2018 1. May]; Available from: https://www.cancer.gov/about-cancer/treatment/types/radiation-therapy/side-effects.
7. Paganetti, H., *Proton therapy physics*. Series in medical physics and biomedical engineering. 2012, Boca Raton, FL: CRC Press. xx, 684 p.
8. Rorvik, E., S. Thornqvist, and K.S. Ytre-Hauge, *The experimental dose ranges influence the LETd dependency of the proton minimum RBE (RBEmin).* Phys Med Biol, 2019. **64**(19): p. 195001.
9. Khan, F. and J. Gibbons, *Khan's The Physics of Radiation Therapy*. 5th ed. 2014: Wolters Kluwer.
10. Cember, H. and T.E. Johnson, *Introduction to health physics*. 4th ed. 2009, New York: McGraw-Hill Medical. xi, 843 p.
11. RADIOLOGY, a.i.I. *4: Radiation Interactions with Tissue*. 2016; Available from: https://radiologykey.com/4-radiation-interactions-with-tissue/.
12. Mohan, R. and D. Grosshans, *Proton therapy - Present and future.* Adv Drug Deliv Rev, 2017. **109**: p. 26-44.
13. Newhauser, W.D. and R. Zhang, *The physics of proton therapy.* Phys Med Biol, 2015. **60**(8): p. R155-209.
14. Levin, W.P., et al., *Proton beam therapy.* Br J Cancer, 2005. **93**(8): p. 849-54.
15. Bethe, H., *Zur theorie des durchgangs schneller korpuskularstrahlen durch materie.* Annalen der Physik, 1930: p. 325-400.
16. Bloch, F., *Zur bremsung rasch bewegter teilchen beim durchgang durch materie.* Annalen der Physik, 1933: p. 285-320.
17. Romano, F., et al., *A Monte Carlo study for the calculation of the average linear energy transfer (LET) distributions for a clinical proton beam line and a radiobiological carbon ion beam line.* Phys Med Biol, 2014. **59**(12): p. 2863-82.

18. International Commission on Radiation, U. and Measurements, *Report 85: Fundamental quantities and units for ionizing radiation.* J ICRU, 2011. **11**(1): p. 1-31.
19. Guan, F., et al., *Analysis of the track- and dose-averaged LET and LET spectra in proton therapy using the geant4 Monte Carlo code.* Med Phys, 2015. **42**(11): p. 6234-47.
20. Wilkens, J.J. and U. Oelfke, *Optimization of radiobiological effects in intensity modulated proton therapy.* Med Phys, 2005. **32**(2): p. 455-65.
21. Mayles, P., A. Nahum, and J.C. Rosenwald, *Handbook of Radiotherapy Physics; Theory and Practice*. 2007: Taylor & Francis Group.
22. Guan, F., et al., *Spatial mapping of the biologic effectiveness of scanned particle beams: towards biologically optimized particle therapy.* Sci Rep, 2015. **5**: p. 9850.
23. Joiner, M. and A.v.d. Kogel, *Basic clinical radiobiology*. 4th ed. 2009, London: Hodder Arnold ;. vi, 375 p.
24. Garcia, L.M., et al., *Fitting the linear-quadratic model to detailed data sets for different dose ranges.* Phys Med Biol, 2006. **51**(11): p. 2813-23.
25. Rorvik, E., et al., *A phenomenological biological dose model for proton therapy based on linear energy transfer spectra.* Med Phys, 2017. **44**(6): p. 2586-2594.
26. Paganetti, H., *Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer.* Phys Med Biol, 2014. **59**(22): p. R419-72.
27. Wedenberg, M. and I. Toma-Dasu, *Disregarding RBE variation in treatment plan comparison may lead to bias in favor of proton plans.* Med Phys, 2014. **41**(9): p. 091706.
28. Wilkens, J.J. and U. Oelfke, *A phenomenological model for the relative biological effectiveness in therapeutic proton beams.* Phys Med Biol, 2004. **49**(13): p. 2811-25.
29. Carabe-Fernandez, A., R.G. Dale, and B. Jones, *The incorporation of the concept of minimum RBE (RbEmin) into the linear-quadratic model and the potential for improved radiobiological analysis of high-LET treatments.* Int J Radiat Biol, 2007. **83**(1): p. 27-39.
30. Dale, R.G. and B. Jones, *The assessment of RBE effects using the concept of biologically effective dose.* Int J Radiat Oncol Biol Phys, 1999. **43**(3): p. 639-45.
31. Rorvik, E., et al., *Exploration and application of phenomenological RBE models for proton therapy.* Phys Med Biol, 2018. **63**(18): p. 185013.
32. Kaufman, L. and P.j. Rousseeuw, *Finding groups in data: An introduction to cluster analysis*. 1990, New York: John Wiley & Sons, inc.
33. Tan, P.-N., M. Steinbach, and V. Kumar, *Introduction to data mining*. 1st ed. 2006, Boston: Pearson Addison Wesley. xxi, 769 p.
34. Fowler, S., R. Roush, and J. Wise, *Concepts of Biology*. 2013, Houston: OpenStax.
35. Unkel, S., C. Belka, and K. Lauber, *On the analysis of clonogenic survival data: Statistical alternatives to the linear-quadratic model.* Radiat Oncol, 2016. **11**: p. 11.
36. Even, A.J.G., et al., *Clustering of multi-parametric functional imaging to identify high-risk subvolumes in non-small cell lung cancer.* Radiother Oncol, 2017. **125**(3): p. 379-384.
37. Datta, N.R., et al., *Efficacy and Safety Evaluation of the Various Therapeutic Options in Locally Advanced Cervix Cancer: A Systematic Review and Network Meta-Analysis of Randomized Clinical Trials.* Int J Radiat Oncol Biol Phys, 2019. **103**(2): p. 411-437.

38. GSI Helmholtzzentrum für Schwerionenforschung. *Particle Irradiation Data Ensemble*. 2019 [cited February 06, 2020; Available from: https://www.gsi.de/en/work/research/biophysics/biophysical_research/radiobiological_modelling/pide_project.htm.
39. Rohatgi, A. *WebPlotDigitizer*. 2019 February 6. 2020]; Available from: https://automeris.io/WebPlotDigitizer/.
40. Baggio, L., et al., *Relative biological effectiveness of light ions in human tumoural cell lines: role of protein p53.* Radiat Prot Dosimetry, 2002. **99**(1-4): p. 211-4.
41. Belli, M., et al., *RBE-LET relationships for cell inactivation and mutation induced by low energy protons in V79 cells: further results at the LNL facility.* Int J Radiat Biol, 1998. **74**(4): p. 501-9.
42. Belli, M., et al., *Inactivation of human normal and tumour cells irradiated with low energy protons.* Int J Radiat Biol, 2000. **76**(6): p. 831-9.
43. Bettega, D., et al., *Inactivation of C3H10T1/2 cells by low energy protons and deuterons.* Int J Radiat Biol, 1998. **73**(3): p. 303-9.
44. Bird, R.P., et al., *Inactivation of synchronized Chinese Hamster V79 cells with charged-particle track segments.* Radiat Res, 1980. **82**(2): p. 277-89.
45. Chaudhary, P., et al., *Relative biological effectiveness variation along monoenergetic and modulated Bragg peaks of a 62-MeV therapeutic proton beam: a preclinical assessment.* Int J Radiat Oncol Biol Phys, 2014. **90**(1): p. 27-35.
46. Folkard, M., et al., *Inactivation of V79 cells by low-energy protons, deuterons and helium-3 ions.* Int J Radiat Biol, 1996. **69**(6): p. 729-38.
47. Folkard, M., et al., *The irradiation of V79 mammalian cells by protons with energies below 2 MeV. Part I: Experimental arrangement and measurements of cell survival.* Int J Radiat Biol, 1989. **56**(3): p. 221-37.
48. Fuhrman Conti, A.M., et al., *Induction of 8-azaguanine resistant mutants in human cultured cells exposed to 31 MeV protons.* Int J Radiat Biol Relat Stud Phys Chem Med, 1988. **53**(3): p. 467-76.
49. Green, L.M., et al., *Response of thyroid follicular cells to gamma irradiation compared to proton irradiation. I. Initial characterization of DNA damage, micronucleus formation, apoptosis, cell survival, and cell cycle phase redistribution.* Radiat Res, 2001. **155**(1 Pt 1): p. 32-42.
50. Green, L.M., et al., *Response of thyroid follicular cells to gamma irradiation compared to proton irradiation: II. The role of connexin 32.* Radiat Res, 2002. **158**(4): p. 475-85.
51. Hei, T.K., et al., *Mutation induction by charged particles of defined linear energy transfer.* Carcinogenesis, 1988. **9**(7): p. 1233-6.
52. Hei, T.K., et al., *Oncogenic transformation by charged particles of defined LET.* Carcinogenesis, 1988. **9**(5): p. 747-50.
53. Howard, M.E., *Caracterization of Relative Biological Effectiveness for Proton Therapy in Human Cancer Cell Lines*, in *Faculty of Mayo Clinic Collage of Medicine*. 2017, Mayo Graduate School.
54. Jeynes, J.C., et al., *"Broadbeam" irradiation of mammalian cells using a vertical microbeam facility.* Radiat Environ Biophys, 2013. **52**(4): p. 513-21.
55. Moertel, H., et al., *Effects of low energy protons on clonogenic survival, DSB repair and cell cycle in human glioblastoma cells and B14 fibroblasts.* Radiother Oncol, 2004. **73 Suppl 2**: p. S115-8.

56. Patel, D., et al., *Optimization of Monte Carlo particle transport parameters and validation of a novel high throughput experimental setup to measure the biological effects of particle beams.* Med Phys, 2017. **44**(11): p. 6061-6073.
57. Perris, A., et al., *Biological effectiveness of low energy protons. I. Survival of Chinese hamster cells.* Int J Radiat Biol Relat Stud Phys Chem Med, 1986. **50**(6): p. 1093-101.
58. Petrovic, I., et al., *Radiobiological analysis of human melanoma cells on the 62 MeV CATANA proton beam.* Int J Radiat Biol, 2006. **82**(4): p. 251-65.
59. Prise, K.M., et al., *The irradiation of V79 mammalian cells by protons with energies below 2 MeV. Part II. Measurement of oxygen enhancement ratios and DNA damage.* Int J Radiat Biol, 1990. **58**(2): p. 261-77.
60. Schettino, G., et al., *Low-dose hypersensitivity in Chinese hamster V79 cells targeted with counted protons using a charged-particle microbeam.* Radiat Res, 2001. **156**(5 Pt 1): p. 526-34.
61. Schuff, J.A., et al., *Relative biological effectiveness measurements of low energy proton and lithium beams on tumor cells.* Nuclear Instruments & Methods in Physics Research Section B-Beam Interactions with Materials and Atoms, 2002. **187**(3): p. 345-353.
62. Sgura, A., et al., *Micronuclei, CREST-positive micronuclei and cell inactivation induced in Chinese hamster cells by radiation with different quality.* Int J Radiat Biol, 2000. **76**(3): p. 367-74.
63. Slonina, D., et al., *Relative biological effectiveness of the 60-MeV therapeutic proton beam at the Institute of Nuclear Physics (IFJ PAN) in Krakow, Poland.* Radiat Environ Biophys, 2014. **53**(4): p. 745-54.
64. Wainson, A.A., et al., *The RBE of accelerated protons in different parts of the Bragg curve.* Br J Radiol, 1972. **45**(535): p. 525-9.
65. Wera, A.C., et al., *Low-LET proton irradiation of A549 non-small cell lung adenocarcinoma cells: dose response and RBE determination.* Radiat Res, 2013. **179**(3): p. 273-81.
66. Mairani, A., et al., *Data-driven RBE parameterization for helium ion beams.* Physics in Medicine and Biology, 2016. **61**(2): p. 888-905.
67. Tilly, N., et al., *The influence of RBE variations in a clinical proton treatment plan for a hypopharynx cancer.* Phys Med Biol, 2005. **50**(12): p. 2765-77.
68. IBM, *IBM SPSS Statistics Algorithms*, IBM, Editor. 2017.
69. Akaike, H., *Citation Classic - a New Look at the Statistical-Model Identification.* Current Contents/Engineering Technology & Applied Sciences, 1981(51): p. 22-22.
70. Burnham, K.P. and D.R. Anderson, *Model Selection and Multimodel Inference: A practical information-theoretic approach*. 2nd ed. 2002: Springer-Verlag.
71. Bacher, J., K. Wenzig, and M. Vogler. *SPSS TwoStep Cluster - a first evaluation*.
72. Burnham, K.P. and D.R. Anderson, *Multimodel inference: understanding AIC and BIC in Model Selection.* Sociological Methods & Research, 2004: p. 261-304.
73. Yang, Y.H., *Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation.* Biometrika, 2005. **92**(4): p. 937-950.
74. de Amorim, R.C. and C. Hennig, *Recovering the number of clusters in data sets with noise features using feature rescaling factors.* Information Sciences, 2015. **324**: p. 126-145.

75. Rousseeuw, P.J., *Silhouettes - a Graphical Aid to the Interpretation and Validation of Cluster-Analysis.* Journal of Computational and Applied Mathematics, 1987. **20**: p. 53-65.
76. Belli, M., A. Campa, and I. Ermolli, *A semi-empirical approach to the evaluation of the relative biological effectiveness of therapeutic proton beams: the methodological framework.* Radiat Res, 1997. **148**(6): p. 592-8.
77. Chen, Y. and S. Ahmad, *Empirical model estimation of relative biological effectiveness for proton beam therapy.* Radiat Prot Dosimetry, 2012. **149**(2): p. 116-23.
78. Carabe, A., et al., *Range uncertainty in proton therapy due to variable biological effectiveness.* Phys Med Biol, 2012. **57**(5): p. 1159-72.
79. Wedenberg, M., B.K. Lind, and B. Hardemark, *A model for the relative biological effectiveness of protons: the tissue specific parameter alpha/beta of photons is a predictor for the sensitivity to LET changes.* Acta Oncol, 2013. **52**(3): p. 580-8.
80. Jones, B., *A Simpler Energy Transfer Efficiency Model to Predict Relative Biological Effect for Protons and Heavier Ions.* Front Oncol, 2015. **5**: p. 184.
81. McNamara, A.L., J. Schuemann, and H. Paganetti, *A phenomenological relative biological effectiveness (RBE) model for proton therapy based on all published in vitro cell survival data.* Phys Med Biol, 2015. **60**(21): p. 8399-416.
82. Mairani, A., et al., *A phenomenological relative biological effectiveness approach for proton therapy based on an improved description of the mixed radiation field.* Phys Med Biol, 2017. **62**(4): p. 1378-1395.
83. Peeler, C., *Assessing the Potential Clinical Impact of Variable Biological Effectiveness in Proton Radiation*. 2016, The University of Texas.
84. MathWorks. *MATLAB R2017b*. 2017 [cited 2017; Available from: https://www.mathworks.com/products/matlab.html.
85. Fjaera, L.F., et al., *Linear energy transfer distributions in the brainstem depending on tumour location in intensity-modulated proton therapy of paediatric cancer.* Acta Oncol, 2017. **56**(6): p. 763-768.
86. Böhlen, T.T., et al., *The FLUKA Code: Developments and Challenges for High Energy and Medical Applications.* Nuclear Data Sheets, 2014. **120**: p. 211 - 214.
87. Ferrari, A., et al., *FLUKA: A multi-particle transport code (Program version 5).* 2005.
88. Vlachoudis, V., *FLAIR: A powerful but user friendly graphical interface for FLUKA.* 2009, Nwe York: Int. Conf. on Mathematics, Computational Methods & Reactor Physics.
89. Henjum, H., *Optimization of proton therapy plans with respect to biological and physical dose distribution*, in *Department of Physics and Technology*. 2018, University of Bergen: Bergen.
90. Frese, M.C., et al., *Application of constant vs. variable relative biological effectiveness in treatment planning of intensity-modulated proton therapy.* Int J Radiat Oncol Biol Phys, 2011. **79**(1): p. 80-8.

Appendix A

Database and database calculations:

In Chapter 3.1 the LET_d in our database was normalized to Cobalt-60 using the values in Table 7.

Table 7: Values used to normalize LET_d values of the database. Where exact values were not listed, interpolating was used to estimate the value.

Reference radiation	$\boldsymbol{LET_d\left[\frac{keV}{\mu m}\right]}$	**Reference**
^{60}Co	0.4	*Howard et al* [53]
^{137}Cs	0.8	
6 MV	0.2	
120 kVp	1.387	*Mairani et al* [66]
200 kVp	1.164	
225 kVp	1.118	
240 kVp	1.092	
250 kVp	1.075	
300 kVp	0.99	

Appendix B

Results from the original cluster analysis where the four final experiments were not excluded.

Table 8: Mean value for each variable considered in the analysis with accompanying standard deviation for each model and for the entire dataset (All clusters), rounded to two decimals. This result includes the later excluded datapoints mentioned in chapter 3.2.

Mean and Standard deviation						
	Cluster 1	***Cluster 2***	***Cluster 3***	***Cluster 4***	***Cluster 5***	***All clusters***
$D_{\min}\,[Gy]$	0.86 ± 0.53	8.02 ± 0	0.66 ± 0.44	0.34 ± 0.17	0.44 ± 0.37	0.75 ± 0.88
$\beta[Gy^{-2}]$	0.06 ± 0.05	0.00 ± 0	0.07 ± 0.07	0.50 ± 0.22	0.05 ± 0.04	0.11 ± 0.17
$LET_d\left[\frac{keV}{\mu m}\right]$	11.59 ± 4.46	1.04 ± 0	2.22 ± 1.74	15.66 ± 3.23	9.33 ± 5.80	0.88 ± 6.20
$\alpha[Gy^{-1}]$	0.33 ± 0.15	0.03 ± 0	0.16 ± 0.09	0.32 ± 0.24	1.07 ± 0.41	0.33 ± 0.32
$\left(\frac{\alpha}{\beta}\right)_x\,[Gy]$	7.54 ± 11.85	17.10 ± 0	4.15 ± 5.25	3.29 ± 0.67	11.19 ± 5.24	6.15 ± 8.45
RBE_{min}	1.17 ± 0.60	0.52 ± 0	1.11 ± 0.29	2.47 ± 0.49	0.79 ± 0.47	1.26 ± 0.64
RBE_{max}	4.51 ± 13.02	1.75 ± 0	2.06 ± 5.39	1.66 ± 2.02	2.49 ± 1.53	2.96 ± 8.66

We observe from Figure 19 that the AIC is lowest at 5 clusters, also in this analysis, and even though 6, 7 and, 8 clusters all gave a slightly better Silhouette score, a 5 cluster model was chosen here as well. This was also the number of clusters chosen automatically by SPSS if no specific number of clusters was decided.

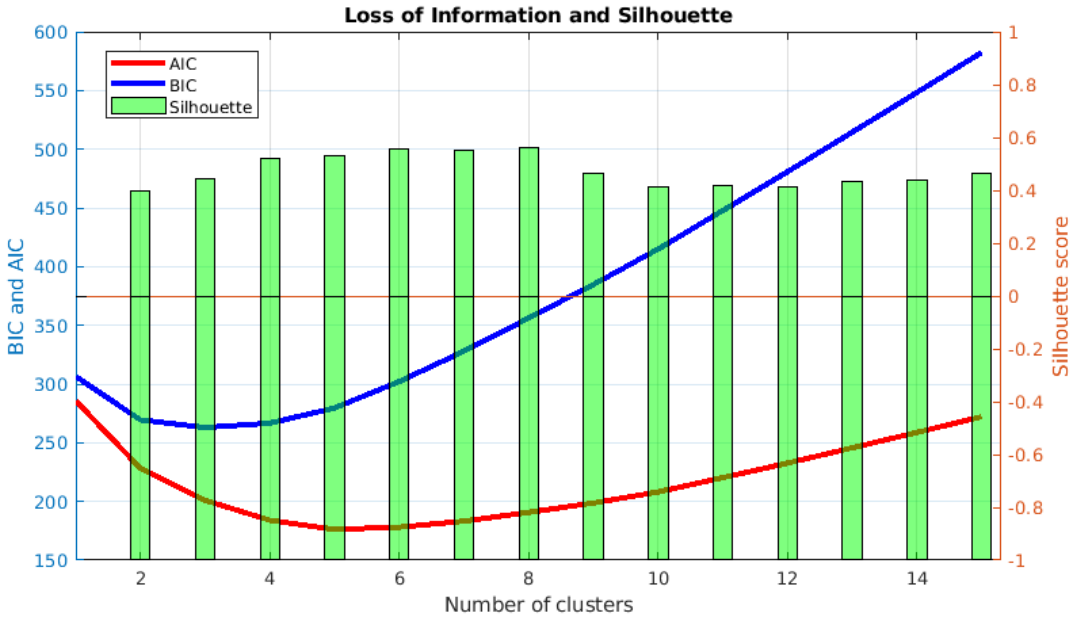

Figure 19: The loss of information as a function of the number of clusters, calculated using Bayesian Information Criterion (Blue) and Akaike Information Criterion (Red), respectively. The Silhouette scores for 2-15 clusters are also depicted as Green bars. This result includes the later excluded datapoints mentioned in chapter 3.2.

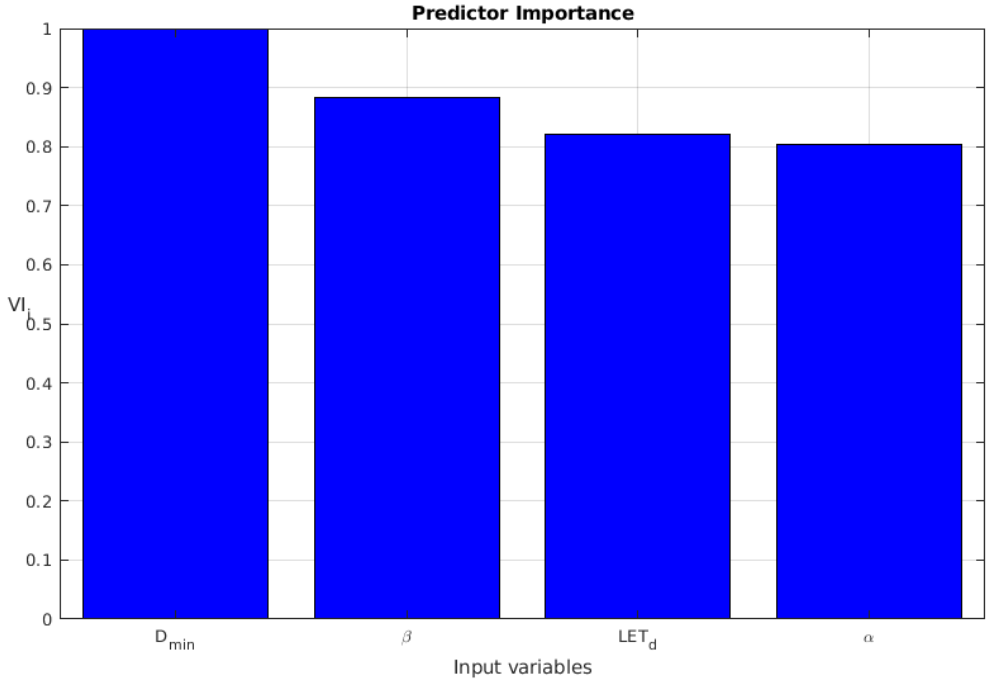

Figure 20: Predictor Importance for each of the input variables, for a 5-cluster model. This result includes the later excluded datapoints mentioned in chapter 3.2.

From Figure 20 we observe a preferable Predictor importance to the one presented in the final analysis. The lowest Predictor importance value was here still over 0.8, for α, and the order was slightly changed as D_{min} and β switched places. The same was observed for LET_d and α.

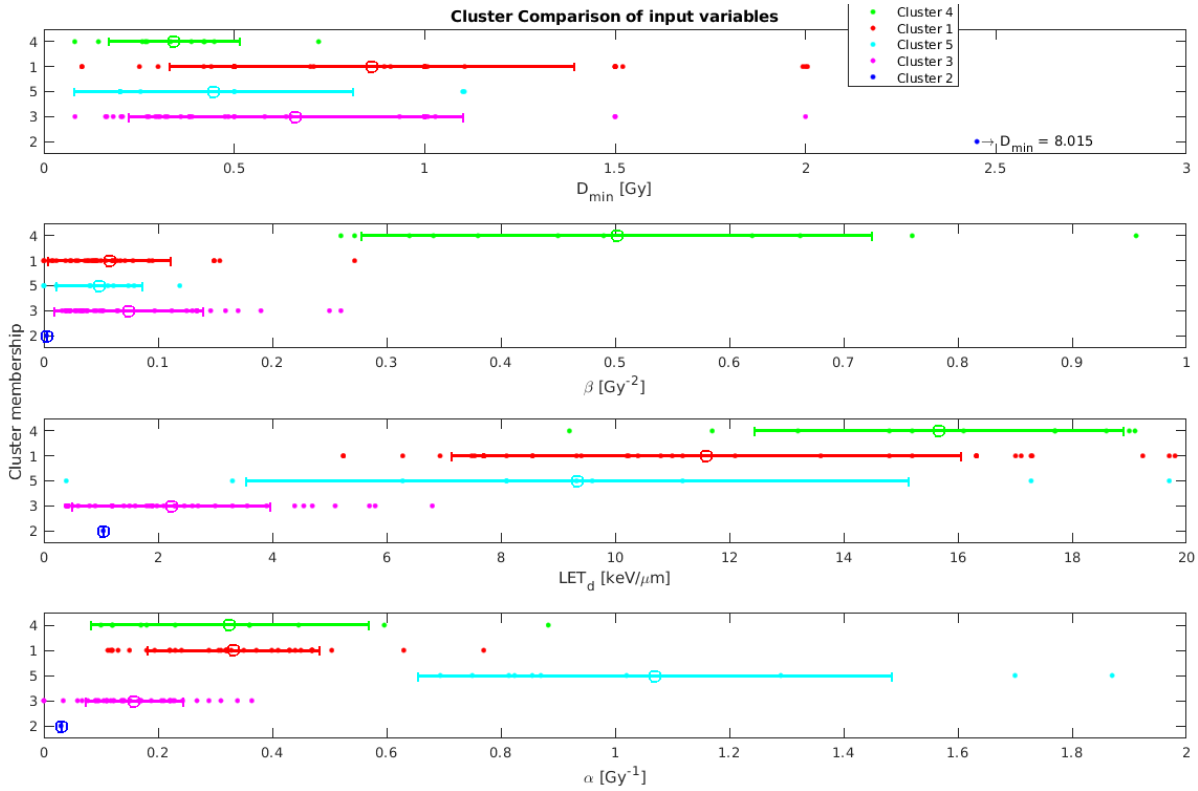

Figure 21: Cluster comparison of input variables. All datapoints in each cluster are shown as dots for each input variable, with mean value marked as a circle with error bars of $\pm$ one standard deviation of the mean. Input variables are listed from top to bottom by decreasing Predictor Importance. This result includes the later excluded datapoints mentioned in chapter 3.2.

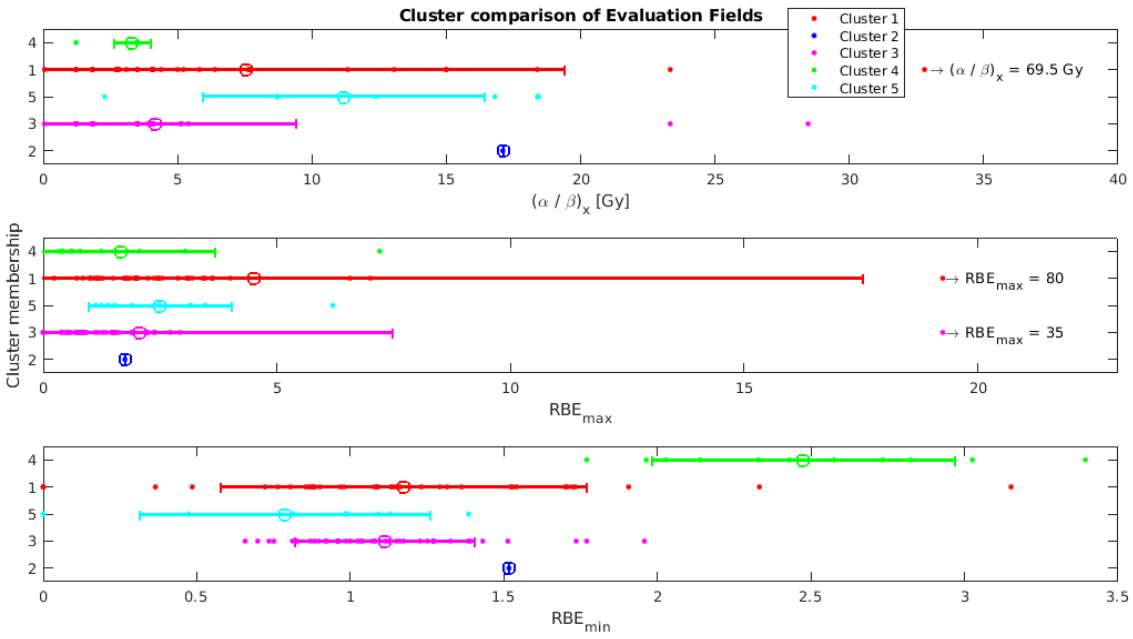

Figure 22: Cluster comparison of evaluation fields. All datapoints in each cluster are shown as dots for each input variable, with mean value marked as a circle with error bars of ± one standard deviation of the mean. Input variables are listed from top to bottom by decreasing Predictor Importance. This result includes the later excluded datapoints mentioned in chapter 3.2.

Appendix C

Model equations: Existing models.

The McNamara model:

The McNamara model can be described by the following equations [31, 81, 83]:

$$RBE_{min}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 1.1012 - 0.0038703Gy^{-\frac{1}{2}}\left(\frac{keV}{\mu m}\right)^{-1}\sqrt{\left(\frac{\alpha}{\beta}\right)_x}\,LET_d \qquad (36)$$

$$RBE_{max}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 0.99064 + \frac{0.35605Gy\left(\frac{keV}{\mu m}\right)^{-1}}{\left(\frac{\alpha}{\beta}\right)_x}LET_d \qquad (37)$$

The Mairani model:

The Mairani model considering only protons can be described by the following equations [31, 82]:

$$RBE_{max,p}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 1 + \frac{0.377Gy\left(\frac{keV}{\mu m}\right)^{-1}}{\left(\frac{\alpha}{\beta}\right)_x}LET_{d,p} \qquad (38)$$

A constant $RBE_{min} = 1$ was assumed for this model.

The Wilkens and Oelfke model:

The Wilkens and Oelfke model can be described by the following equation [28, 83, 90]:

$$RBE_{max}(\lambda, \alpha_x, LET_d) = 1 + \frac{\lambda}{\alpha_x}\left(LET_d - 0.5\frac{keV}{\mu m}\right) \quad (39)$$

Where λ is a linear parameter fitting the equation to experimental data. By assuming $\alpha_x = 0.112Gy^{-1}$ and $\beta_x = 0.0298Gy^{-2}$, which was assumed in the paper that suggested the model, the resulting model equation can be written as [31]:

$$RBE_{max}(LET_d) = 0.892 + 0.179\left(\frac{keV}{\mu m}\right)^{-1} LET_d \quad (40)$$

A constant $RBE_{min} = 1$ was assumed for this model.

The Peeler model:

The Peeler model can be described by the following equations [31, 83]:

$$RBE_{max}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 0.75 + \frac{0.00143Gy\left(\frac{keV}{\mu m}\right)^{-3}}{\left(\frac{\alpha}{\beta}\right)_x} LET_d^3 \quad (41)$$

$$RBE_{min}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 1.24 + 0.00074Gy^{-1}\left(\frac{keV}{\mu m}\right)^{-3}\left(\frac{\alpha}{\beta}\right)_x LET_d^3 \quad (42)$$

The Wedenberg model:

The Wedenberg model can be described by the following equation [31, 79, 83]:

$$RBE_{max}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 1 + \frac{0.434Gy\left(\frac{keV}{\mu m}\right)^{-1}}{\left(\frac{\alpha}{\beta}\right)_x} LET_d \quad (43)$$

By restricting the database, and thus limiting the LET_d to $20\frac{keV}{\mu m}$, the fitting variable is reduced to $0.378Gy\left(\frac{keV}{\mu m}\right)^{-1}$.

A constant $RBE_{min} = 1$ was assumed for this model.

The unweighted linear Rørvik model:

The unweighted linear Rørvik model can be described by the following equation [25, 31]:

$$RBE_{max}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 1 + \frac{0.645Gy\left(\frac{keV}{\mu m}\right)^{-1}}{\left(\frac{\alpha}{\beta}\right)_x} LET_d \qquad (44)$$

A constant $RBE_{min} = 1$ was assumed for this model.

The Tilly models:

The model equation for RBE_{max} for the $\left(\frac{\alpha}{\beta}\right)_x = 2\ Gy$ model is given as [31, 67]:

$$RBE_{max}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 1 + \frac{0.309Gy\left(\frac{keV}{\mu m}\right)^{-1}}{\left(\frac{\alpha}{\beta}\right)_x} LET_d \qquad (45)$$

And the model equation for RBE_{max} for the $\left(\frac{\alpha}{\beta}\right)_x = 10\ Gy$ model is given as:

$$RBE_{max}\left(LET_d, \left(\frac{\alpha}{\beta}\right)_x\right) = 1 + \frac{0.550964Gy\left(\frac{keV}{\mu m}\right)^{-1}}{\left(\frac{\alpha}{\beta}\right)_x} LET_d \qquad (46)$$

A constant $RBE_{min} = 1$ was assumed for both models.

Appendix D

A short description of the scripts used in Chapter 3.5.

Table 9: Description of the scripts used in the FLUKA MC simulation process.

Script	Description	Provided by
Sort_dicom.py	Converts the DICOM files from eclipse to readable files for FLUKA.	Lars Fredrik Fjæra
Convert_to_dicom.py	Converts the output files from the FLUKA MC simulations into DICOM files where dose, LET and RBE distribution are calculated.	
Plot_dicom.py	Plots the DICOM files from the *convert_to_dicom.py* script into 2D plots.	
Plot_1d_dicom.py	Plots the DICOM files from the *convert_to_dicom.py* script into 1D plots	
Dose_RT_to_DVH_1.1.py	Converts the information from the DICOM files from the *convert_to_dicom.py* script into text files with DVH information.	Helge Henjum
Plot_DVH.py	Plots the DVHs based on the information in the text files from the *Dose_RT_to_DVH_1.1.py* script.	

Appendix E

FLUKA simulation and dose calculations

The LET_d was increasing towards the distal edge of the SOBP.

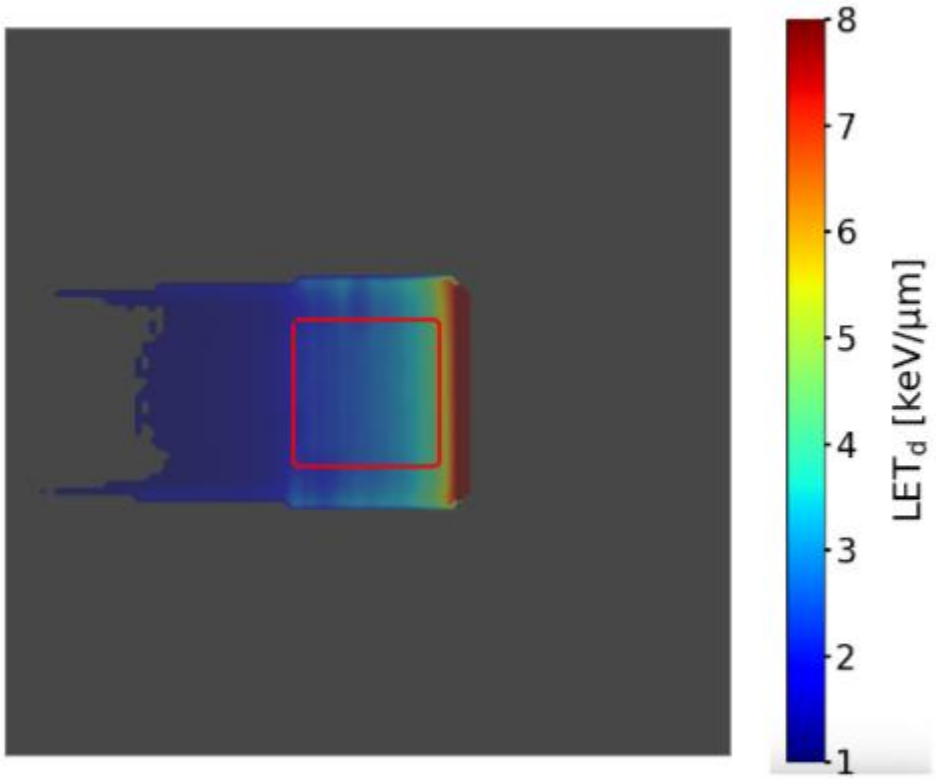

Figure 23: LET_d distribution based on the scoring file created in the FLUKA simulation.